PRINT ISSN : 2395-6011
ONLINE ISSN : 2395-602X

TECHNO SCIENCE ACADEMY
The International Open Access Publisher

Scientific Journal Impact Factor : 7.214

D Y PATIL GROUP
Dr D Y Patil School of Engineering, Lohegaon, Pune

IJSRST

FIRST INTERNATIONAL CONFERENCE ON COMPUTER ENGINEERING DYP-ICCE 2020

Organised by

Dr. D. Y. Patil School of Engineering
Dr Ajeenkya DY Patil Knowledge City Lohegaon, Pune,
Maharashtra, India
Department of Computer Engineering and Association of
Computer Engineering in Association with
International Journal of Scientific Research in Science and Technology
Sponsored By
TechnoScience Academy (The International open Access Publisher)

VOLUME 5, ISSUE 8, NOVEMBER-DECEMBER-2020

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN SCIENCE & TECHNOLOGY

Peer-Reviewed International Scientific Refereed Research

Email : editor@ijsrst.com Website : http://ijsrst.com

ace
association of computer engineering
recoding engineers!

First International Conference on Computer Engineering

DYP-ICCE 2020

17th - 18th December 2020

In Association With

International Journal of Scientific Research in Science and Technology

Print ISSN: 2395-6011 Online ISSN : 2395-602X

Volume 5, Issue 8, November-December-2020

International Peer Reviewed, Open Access Journal

Organised by

Department of Computer Engineering, Dr. D. Y. Patil School of Engineering
Dr Ajeenkya DY Patil Knowledge City Lohegaon, Pune, Maharashtra, India
and Association of Computer Engineering
in Association with
International Journal of Scientific Research in Science and Technology

Sponsored By
TechnoScience Academy (The International open Access Publisher)

Published By
[www.technoscienceacademy.com]

CHIEF PATRON
Dr. Ajeenkya D Y Patil
Chairman & Chancellor, Ajeenkya D Y Patil University, Lohegaon, Pune

PATRON
Dr. E. B. Khedkar
Vice President, Ajinkya D Y Patil Group

Dr. Sushant Patil
Advisor, Dr D Y Patil Technical Campus, Lohegaon, Pune

Dr. Ashok Kasnale
Principal, Dr. D Y Patil School of Engineering

Dr. Pankaj Agarkar
HOD, Computer Engineering, Dr. D Y Patil School of Engineering

CONVENER
Prof Monika Dangore
Assistant Professor, Computer Engineering, Dr. D Y Patil School of Engineering

CO-CONVENER
Prof Pooja Shinde
Assistant Professor, Computer Engineering, Dr. D Y Patil School of Engineering

Session Chair

Dr. Sheshang D. Degadwala
Head of Computer Department, Sigma Institute of Engineering, Vadodara, Gujarat, India

Mr Ashwath Salimath
Data Scientist, SSE Airtricity Dublin, Ireland

Mr Darshan Rathi
Senior Software Engineer, Marsh and McClenen Global services India Pvt Ltd. Pune

Dr. Manoj S. Kathane
Principal, Thakur Shivkumarsingh Memorial Engineering College Burhanpur, M.P.

Mr Ankit Kumar Verma
Team lead- Infonyx Pyt. Ltd.Pune

Mr Sagar Pradip Patil
Senior Information Security Analyst at Phreesia. Inc Baner, Pune

Prof Ati Jain
Assistant Professor Sage University Indore, M.P.

Mr Aditya Prakash,
Operational Manager (Finance), Think and Learn Pvt ltd., Bangalore

Dr Shilpa Gite,
Assistant Professor, Computer Science, Symbiosis Institute of Technology, Pune

Mr Pandurang Humbe
Associate Consultant ZS Associates WTC Kharadi Pune

Ms Joyce Julius Silvera
Tesla, Inc. Palo Alto, Silicon Valley

Ms Divya Saxena
Senior Software Testing Engineer, Nihilent Limited, Pune

Mr Somik Sutradhar
Indian Railways/ JE/IT

Mr Dhananjay Khade
Test Engineer , Infosys ltd , Pune

Ms Jyotsna Patil
Devops Engineer, Cybage Pune

DYPSOE Session Coordinators

Prof Pooja Shinde

Dr Sunil Rathod

Dr Sunil Rathod

Prof Pallavi Shimpi

Prof Yogesh Mali

Prof Vandana Chavhan

Prof Vishal Borate

Prof Ashwini Pandagale

Prof Ajita Mahapadi

Prof Prajakta Khairnar

Prof Jayashree Chaudhari

Prof Monika Dangore

Prof Chaitanya Bhosale

Prof Amruta Chitari

Prof Nilesh Mali

Prof Niyamat Ujloomwale

Prof Prashant Mandale

Prof Faraz Bagwan

Conference Report

Computer Engineering Department of Dr D Y Patil School of Engineering, Lohegaon, Pune had organized the First International Conference on Computer Engineering "DYPICCE 2020" on 17th-18th December 2020. This Conference was organized in association with International Journal of Scientific Research in Science and Technology (IJSRST) and was sponsored by Techno Science Academy-The International Open Access Publisher.

For this conference, International Entries for Paper Submission were received from South Africa, Egypt and Kenya. At National Level, entries were received from Madhya Pradesh, Rajasthan and Maharashtra.

The Two Days E-Conference was organized in to 16 Sessions to be conducted on Zoom Platform. Each Session was organized by a Session Coordinator (Internal Faculty of Computer Department) and judged by a Session Chair. The conference had received 16 Session Chairs who are either Experienced IT Professionals or Senior Academicians working at International Places like U.S, Ireland and at National Level, at Madhya Pradesh, Karnataka, Maharashtra and Gujrat.

On Day-I (17th December), Conference Inauguration and Presentation by Winners of Best Paper Award was conducted from 10.00 AM-12.00 PM under Dr. Sheshang D. Degadwala-Executive Editor (IJSRST) as Session Chair. On Day-II, 15 Sessions were conducted in parallel from 2.00 PM to 4.00 PM.

The conference was executed successfully under the support and guidance of Dr Pankaj Agarkar-HOD Computer Engineering. Dr Ashok Kasnale- Principal DYPSOE gave huge support and encouragement to the Conference. Dr Farook Sayyad- Dean Academics-DYPSOE motivated and encouraged the conference. The conference was successfully coordinated by Convener Prof Monika Dangore and Co-Convener Prof Pooja Shinde along with Dr Bhavesh Kataria - Executive Editor (IJSRST).

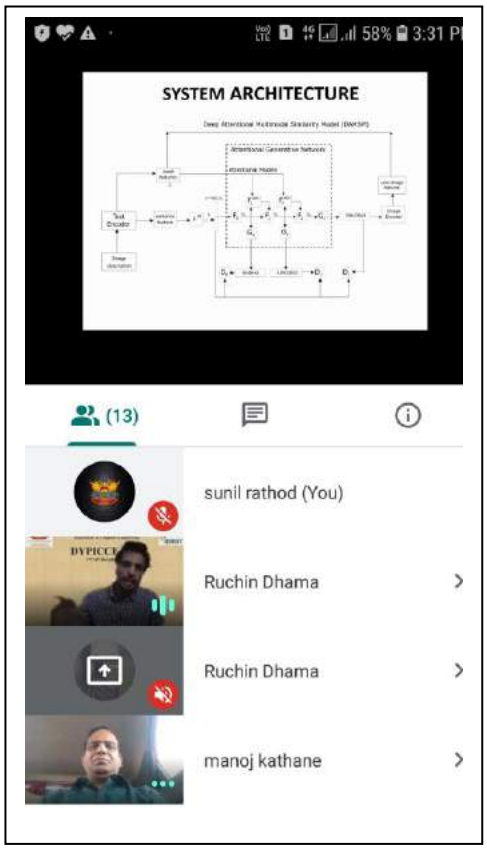
Day-I : 17th December 2020

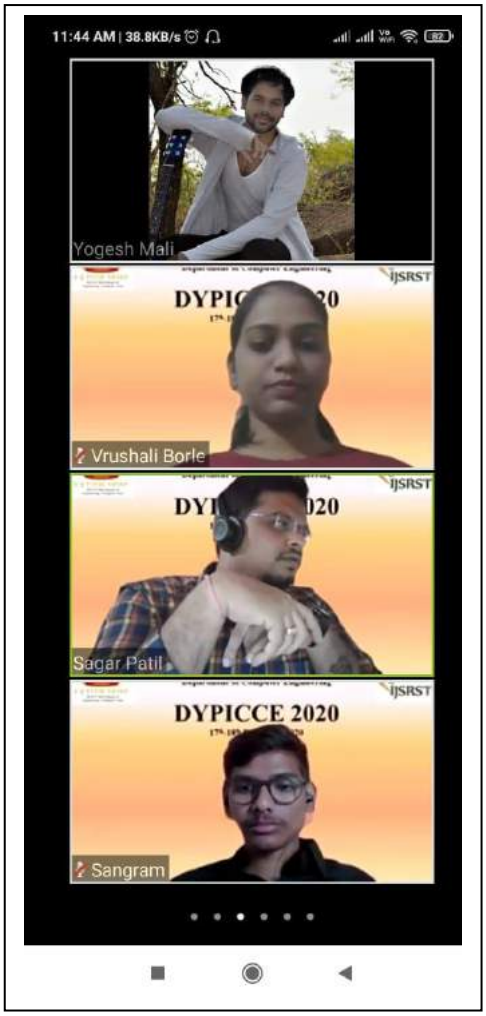
Winners of Best Papers Award (declared by IJSRST) :

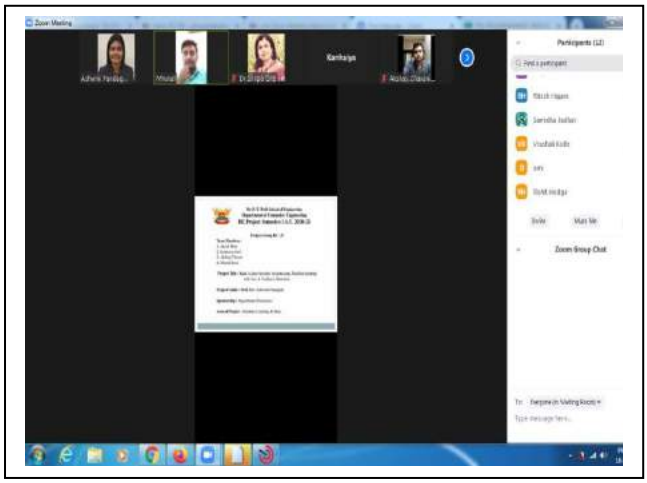
Rank	Paper Title	Name of Authors
1	Development Of Tumorhunt Algorithm For Brain Tumor Segmentation Using Machine Learning Cnn Model	Smita Kakade (JJTU Rajasthan), Deepali Hirolikar (JJTU Rajasthan), Dr.Vinod M. Vaze(JJTU Rajasthan), Dr. Pankaj Agarkar (DYPSOE, Pune)
2	Artificial Intelligence based COVID-19 classification by using Deep Learning and Convolutional Neural Network	Omkar Gaikwad, Divyanshu Tripathi, Madhuri Dange , Prof.Pallavi Shimpi (Dr.D.Y.Patil School Of Engineering, Pune)
3	A Framework for Analyzing Real-Time Tweets to Detect Terrorist Activities	Akshay Karale ,Pranav Shinde ,Pushpak Patil ,Sanjay Parmar ,Prof. Niyamat Ujloomwale (Dr.D.Y.Patil School Of Engineering, Pune)

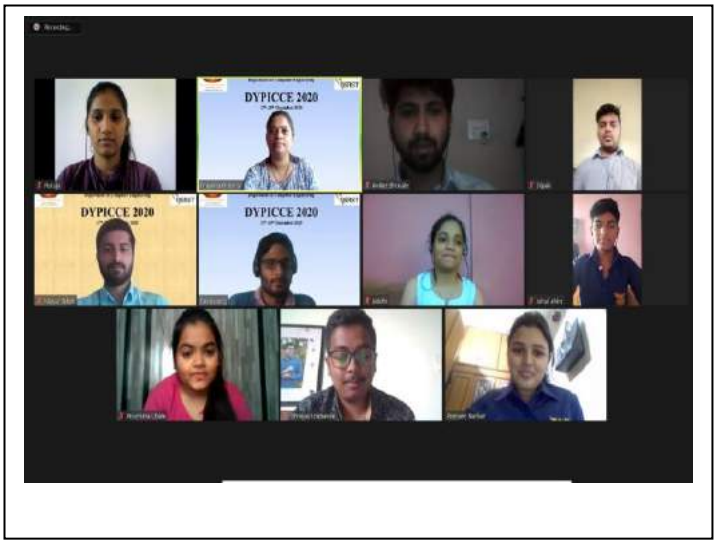
Inaugural Session

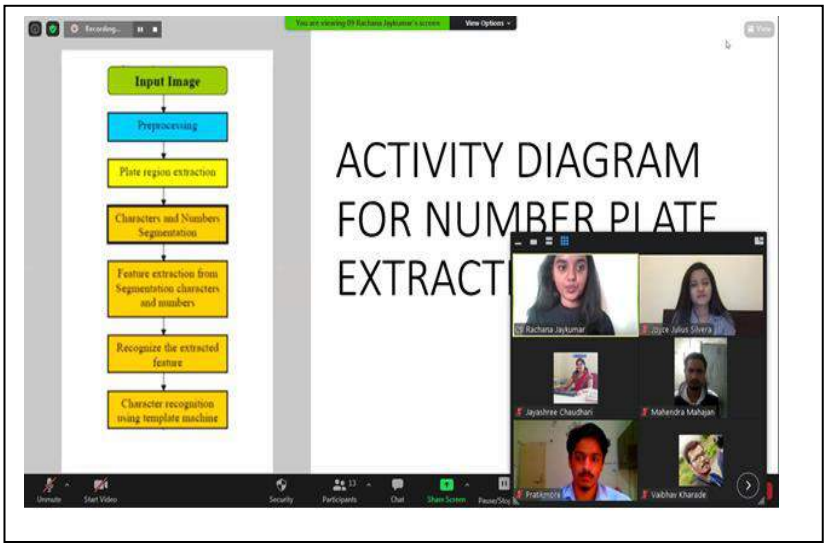
Day-II : 18th December 2020



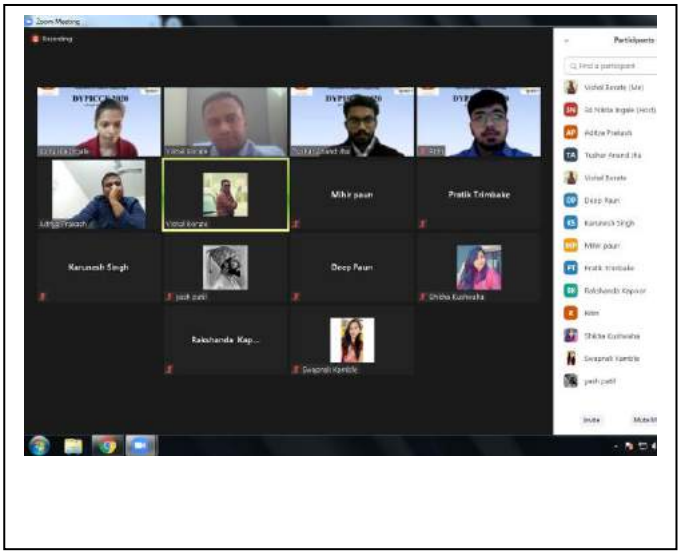


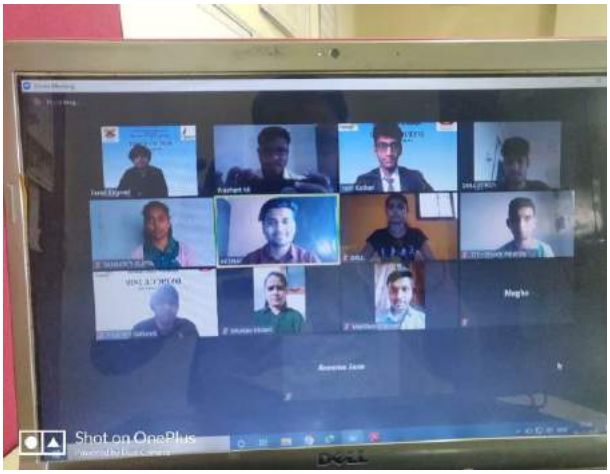


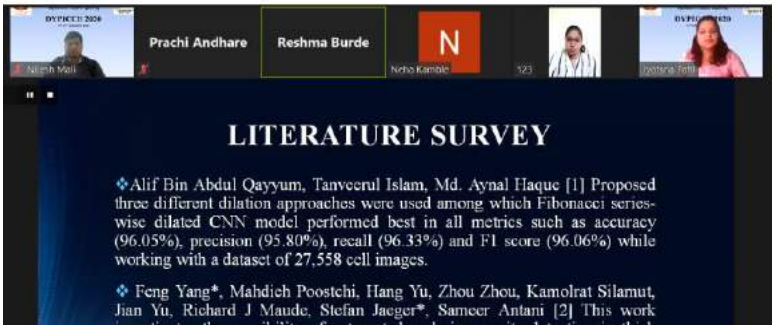

DYPICCE 2020
DYPICCE 2020
DYPICCE 2020

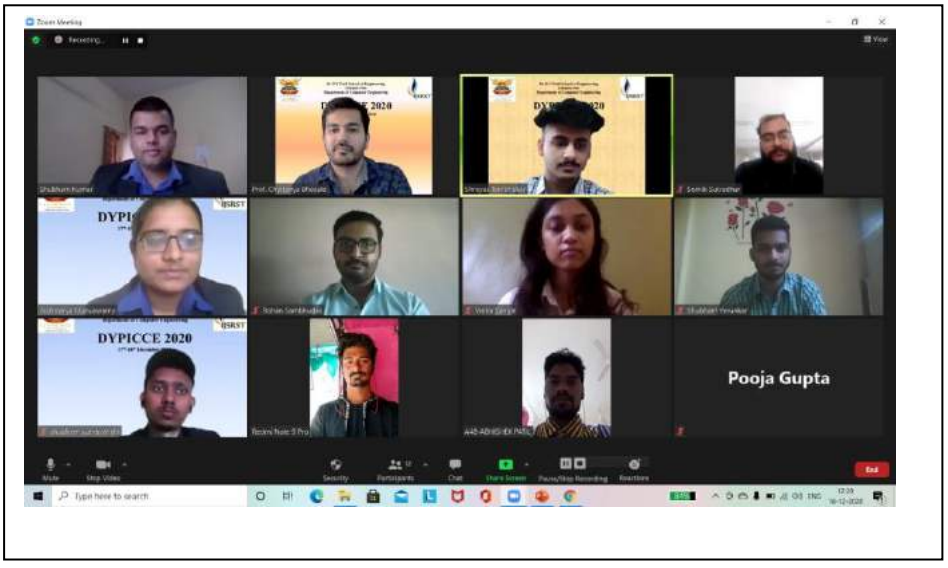

Input Image
Preprocessing
Plate region extraction
Characters and Numbers Segmentation
Feature extraction from Segmentation characters and numbers
Recognize the extracted feature
Character recognition using template machine
ACTIVITY DIAGRAM
FOR NUMBER PLATE
EXTRACT


MOTIVATION:-
Komal


Mihir paun
Pratik Trimbake
Karunesh Singh
Deep Paun
Dhiba Kushwaha
Rakshanda Kap...


Shot on OnePlus


Prachi Andhare
Reshma Burde
LITERATURE SURVEY
Alif Bin Abdul Qayyum, Tanveerul Islam, Md. Aynal Haque [1] Proposed three different dilation approaches were used among which Fibonacci series-wise dilated CNN model performed best in all metrics such as accuracy (96.05%), precision (95.80%), recall (96.33%) and F1 score (96.06%) while working with a dataset of 27,558 cell images.
Feng Yang*, Mahdieh Poostchi, Hang Yu, Zhou Zhou, Kamolrat Silamut, Jian Yu, Richard J Maude, Stefan Jaeger*, Sameer Antani [2] This work


DYPICCE 2020
Pooja Gupta

CONTENTS

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Holistic Empirical Study on Cloud Computing Ecosystem Security Issues

Peter S. Nyakomitta[1*], Silvance O. Abeka[2], Kwach Johnson Kisera[3]
[1*]Faculty of Biological and Physical Sciences, Tom Mboya University College, P.O. Box 199-40300, Homa Bay, Kenya
[2]Department of School of Informatics and Innovative Systems, Jaramogi Oginga Odinga University of Science & Technology, Bondo, Kenya
[3]Faculty of Biological and Physical Sciences, Tom Mboya University College, P.O. Box 199-40300, Homa Bay, Kenya

ABSTRACT

The advent of cloud computing technology offers gorgeous and innovative computing services through resource pooling and virtualization techniques. Cloud providers deliver various types of computing services through the deployment delivery models to customers according to a pay-per-use economic model. However, this technology shift introduces a new concern for enterprises and businesses regarding their data security issues. The paper intends to provide holistic empirically study of cloud computing ecosystem security issues focusing on the current state-of-the-art of cloud computing delivery, deployment models, cloud virtulization technology and network component. The study deployed systematic empirical study approach that was used to review the most profound literature on cloud computing ecosystem in-line with security issues prompted by the fact that service provider necessarily has access to all the data on the cloud and can accidentally or deliberately disclose it or use it for unauthorized purposes. These findings can be used to understand the potential security issues on cloud computing ecosystem hence the study proposed a model with security components to be embedded in cloud based system. It includes authentication and authorization mechanism as a check for the identity of both cloud subscribers and providers.

Keywords : Cloud Computing Ecosystem , Virtualization, Network Isuue, Service Model, Deployment Model

I. INTRODUCTION

Cloud computing technology is a new paradigm that offers the next generation with internet-based, highly scalable distributed computing systems in which computational resources are offered 'as a service'. The most widely used definition of the cloud computing technology was introduced by [1] under NIST as "*a technology for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.*". The two main key characteristics of cloud computing models are Multi-tenancy and elasticity. Multi-Tenancy enables sharing the same service instance among different tenants while elasticity is the ability to right-size resources as they are needed. It allows optimization of system and captures all transactions. Fiqure 1, depicts cloud computing ecosystem as discussed in the sub-sections below.

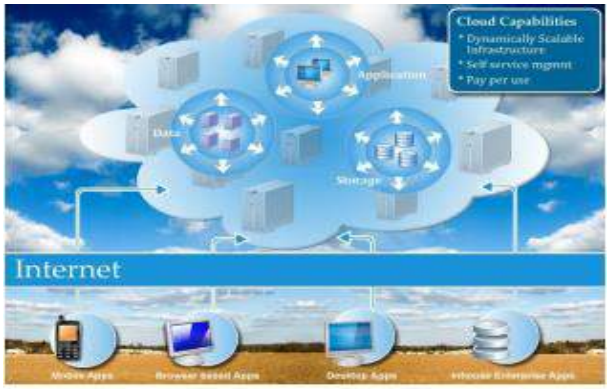


Figure 1 : Conceptual view of cloud computing

Some of the services delivered in cloud computing includes: applications, support services, mail-filtering services, storage services etc as indicated in fiqure 1. The cloud model has motivated industry and academia to adopt cloud computing to host a wide spectrum of applications ranging from high computationally intensive applications down to light weight services [2]. But as more and more information on individuals and companies are placed in the cloud, concerns are beginning to grow about just how safe an environment it is. Despite of all the hype surrounding the cloud, customers are still reluctant to deploy their business in the cloud. Security issues in cloud computing has played a major hindrance as indicated by [3] in slowing down its adoption, infact ranked first as the greatest challenge issue of cloud computing as depicted in figure 2.

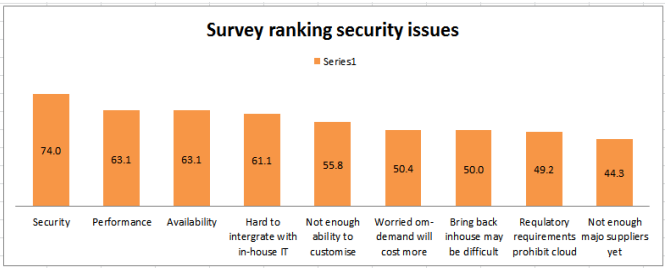


Figure 2 : IDC survey ranking security issues, 2008

From one point of view, security could improve due to centralization of data and increased security-focused resources. According to [4], assessing the quality of cloud providers' approach to security is difficult because many cloud providers will not expose their infrastructure to customers. The current acceptance of cloud computing is linked with various challenges because users are still skeptical about its authenticity. The cloud act as a big black box, nothing inside the cloud is visible to the clients and clients have no idea or control over what happens inside a cloud even if the cloud provider is honest, its only malicious system admins who can tamper with virtual machines and violate confidentiality and integrity.[5]. In this paper, we conducted holistic empirical study on cloud computing ecosystem security issues that has emanated in the cloud environment, laying emphasis on service delivery model, deployment model and the network section.

II. Research Objectives

The objective of the study was to carry out an holistic empirical review of the security issues in cloud computing ecosystem specifically on delivery model, service model, other technologies like virtulization and network and propose a security model that can be used to eliminate ilicit access to the cloud.

CURRENT STATE-OF-THE-ART OF CLOUD COMPUTING ECOSYSTEM

The sub-section presents the review of cloud computing issues in regards to cloud computing deployment model, delivery service, network issues and vitualization technology.

2.1. Deployment Models and its security issues:

Based on the requirements and the services provided by the companies to the subscribers, cloud computing can be deployed in an organization through several deployment models. The sub-section below details the most common types of cloud deployments with their security limitations.

2.1.1. Private cloud.

Private cloud model is generally deployed within an organization and is limited only for the internal access by individuals of that organization. Sub-section below details some specific security issues towards this private cloud model.

Elastic Perimeter- A cloud infrastructure, particularly comprising of private cloud, creates an elastic perimeter. Various departments and users throughout the organization allow sharing of different resources to increase facility of access but unfortunately lead to data breach problem. Moreover,[6] states that elasticity of various cloud based resources would lead to store replicated data on untrusted hosts and this would then lead to enormous risks to data privacy.

Security Control:- The organizations those who are using private cloud infrastructure should need to ensure that effective control of the new environment. [7] stated that the private cloud management architecture should enable management to view security aspects of the environment and show the current threat levels to the organization. The control oversight is to be provided through a web based dashboard that translates the security issues into understandable languages.

2.1.2. Public cloud

Public cloud model is employed by the organization for gaining access to various resources, web applications, and services over any of internet, intranet as well as extranet. Public clouds providers are large targets for hackers. Sub-section below represents security issues related to public cloud model.

Cloning and Resource Pooling:- Cloning deals with replicating or duplicating of the data. According to [8], cloning leads to data leakage problems revealing the machine's authenticity. While [9] describes resource pooling as a service provided to the users by the provider to use various resources and share the same according to their application demand. Resource Pooling relates to the unauthorized access due to sharing through the same network. While the study on Virtual and Cloud Computing by various researches states that a virtual machine can easily be provisioned, they can also be inversed to previous cases, paused, easily restarted, readily cloned and migrated between two physical servers, leading to non-auditable security threats Motility.

Motility of Data and Data residuals:- For the best use of resources, data often is moved to cloud infrastructure. As a result the enterprise would be devoid of the location where data is put on the cloud. This is true with public cloud. With this data movement, the residuals of data is left behind which may be accessed by unauthorized users. According to [10] data-remnant causes very less security threats in private cloud but severe security issues may evolve in public cloud donations. This again may lead to data security threats like data leakage, data remnants and inconsistent data, as stated by [11]. The authors have also mentioned that in order to solve the problems with data storage the optimal solution of cryptography can be thought of effectively.

Shared Multi-tenant Environment:- Multi-tenancy is one of the very vital attribute of cloud computing, which allows multiple users to run their distinct applications concurrently on the same physical infrastructure hiding user data from each other [12]. But the shared multi-tenant character of public cloud adds security risks such as illegal access of data by other renter using the same hardware. A multi-tenant environment might also depict some resource contention issues when any tenant consumes some unequal amount of resources. This might be either

due to genuine periodic requirements or any hack attack [13], has shown that multi-tenancy makes the impact of VM Hopping attack potentially larger than conventional IT environment.

2.1.3. Hybrid clouds.

Hybrid cloud is the combination of two or more clouds (public and/or private). It is an environment providing multiple service suppliers, both internal and external. A hybrid cloud can be considered an intermediate stage as it capitalizes on the benefits of both public and private cloud. But hybrid cloud isn't perfect; it still includes a few security obstacles [14] . The sub-section below discusses some of the security limitations to hybrid cloud.

Lack of Encryption:- Data encryption is a process that helps to address various external and malicious threats. Unencrypted data is vulnerable for susceptible data, as it does not provide any security mechanism. These unencrypted data can easily be accessed by unauthorized users. According to [15], unencrypted data risks the user data leading cloud server to escape various data information to unauthorized users. These unencrypted, insecure data, as per [16], incite the malicious users to misuse the data one or the other way.

Absence of data redundancy:- Problems are inevitable for any cloud providers even though they took best efforts. Hybrid cloud is a complex system. That management has limited experience in managing and that creates great risk. Cloud architects need redundancy across data centers to moderate the impact of an outage in a single data center. A lack of redundancy can become a serious security risk in hybrid cloud, specifically if redundant copies of data are not distributed across data centers. It's easier to move virtual machine (VM) instances between data centers than between large data sets.

Compliance:-In a hybrid cloud maintaining and demonstrating compliance is more difficult. Not only you have to ensure that your public cloud provider and private cloud are in compliance, but also must demonstrate that the means of coordination between the two clouds is compliant. For example if your company works with payment card data, you may be able to demonstrate that both your internal systems and your cloud provider are compliant with the Payment Card Industry Data Security Standard (PCI DSS).

Risk management:- Information security is very difficult to manage risk from a business perspective. Cloud computing (hybrid cloud in particular) uses new application programming interfaces (APIs), requires complex network configurations, and pushes the limits of traditional system administrators' knowledge and abilities.These factors introduce new types of threats.

2.2. Delivery Service models and its security issues.

According to [17], in "The NIST Definition of Cloud Computing," Special Publication 800–145, NIST, broadly divides cloud delivery services into three service models as described in the sub-section below:

2.2.1. Software as a Service (SaaS):

A service is classified as a software if it allows the end user to access and use a provider software application that is hosted, deployed, and managed by the provider from various devices through a thin client interface such as web browser. The users normally have limited control over the application, and are restricted in how they can use and interact with the application. The burden of the security lies with the cloud provider. In part, this is because of the degree of integrated

functionality with minimal user control or extensibility. Sub-section below discusses the various security issues with SaaS.

Application security- SaaS applications are typically delivered via the Internet through a Web browser. However, flaws in web applications may create vulnerabilities for the SaaS aplications. Attackers have been using the web to compromise user's computers and perform malicious activities such as steal sensitive data [18]. Security challenges in SaaS applications are not different from any web application technology, but traditional security solutions do not effectively protect it from attacks, so new approaches are necessary [19].

Multi-tenancy- According to [20], SaaS applications can be grouped into maturity models that are determined by the characteristics such as: scalability, configurability via metadata, and multi-tenancy. In the first maturity model, each customer has his own customized instance of the software. This model has drawbacks since data from multiple tenants is likely to be stored in the same database, the risk of data leakage between these tenants is high.

Data security:- Data security is a common concern for any technology, but it becomes a major challenge when SaaS users have to rely on their providers for proper security. In SaaS, organizational data is often processed in plaintext and stored in the cloud [21] states that the SaaS provider is the one responsible for the security of the data while is being processed and stored, also data backup is a critical aspect in order to facilitate recovery in case of disaster, but it introduces security concerns as well [22]. Cloud providers can subcontract other services such as backup from third-party service providers, which may raise concerns. In the world of SaaS, the process of compliance is complex because data is located in the provider's datacenters, which may introduce regulatory compliance issues such as data privacy, segregation and security that must be enforced by the provider.

Accessibility:- Accessing applications over the internet via web browser makes access from any network device easier, including public computers and mobile devices. However, it also exposes the service to additional security risks. The [23] has released a document that describes the current state of mobile computing and the top threats in this area such as information stealing mobile malware, insecure networks (WiFi), vulnerabilities found in the device OS and official applications, insecure marketplaces, and proximity-based hacking.

2.2.2. Platform-as-a-Service. (PaaS)

PaaS is a moder in which a layer of software or development environment is encapsulated and offered as aservice, upon which other higher levels of services are built. The cloud user has the freedom to build his own applications, which run on the providers infrastructure. PaaS application security comprises two software layers: According [24], security of the PaaS platform itself (i.e., runtime engine), and Security of customer applications deployed on a PaaS platform. PaaS providers are responsible for securing the platform software stack that includes the runtime engine that runs the customer applications. PaaS data security issues are described as follows:

SOA related security issues – the PaaS model is based on the Service-oriented Architecture (SOA) model. This leads to inheriting all security issues that exist in the SOA domain such as DOS attacks, Man-in-the-middle attacks, XML-related attacks, Replay attacks, Dictionary attacks, Injection attacks and input validation related attacks [25]. Mutual authentication, authorization and WS-Security standards are important to secure the cloud provided services. This

security issue is a shared responsibility among cloud providers, service providers and consumers.

API Security - PaaS may offer APIs that deliver management functions such as business functions, security functions, application management etc. Such APIs should be provided with security controls and standards implemented, such as [29]. to enforce consistent authentication and authorization on calls to such APIs. Moreover, there is a need for the isolation of APIs in memory. This issue is under the responsibility of the cloud service provider.

Third-party relationships:- PaaS does not only provide traditional programming languages, but also does it offer third-party web services components such as mashups [27]. Mashups combine more than one source element into a single integrated unit. Thus, PaaS models also inherit security issues related to mashups such as data and network security [28]. Also, PaaS users have to depend on both the security of web-hosted development tools and third-party services.

Underlying infrastructure security:- In PaaS, developers do not usually have access to the underlying layers, so providers are responsible for securing the underlying infrastructure as well as the applications services [29]. Even when developers are in control of the security of their applications, they do not have the assurance that the development environment tools provided by a PaaS provider are secure.

2.2.3. Infrastructure-as-a-Service. (IaaS)

It's a model that provides infrastructure components to clients on demand. The infrastructure components include virtual machines, storage, networks, firewalls and so. With the Iaas, clients have direct access to the lowest-level software in the stack that is operating system which are exposed to some cloud security risk. Sub-section below presents the security issues with IaaS.

Data Leakage Protection and Usage Monitoring:- Data stored in IaaS infrastructure in both private and public cloud needs to be monitored closely [30]. This is essential when IaaS is deployed in public cloud. In this, it should be known that who is accessing the information, how it is accessed, location from where it is accessed and what happened to accessed information later. These problems can be solved by using modern Rights Management services applying restriction to business critical data. Policies for information need to be created and deployed. In addition, transparent process can be created that monitors information usage.

End to End Logging and Reporting:- The effective deployment of IaaS demands comprehensive logging and reporting in place. Robust logging and reporting solutions helps to keep track of where the information is, who accesses it, which machines are handling it and which storage arrays are responsible for it. These solutions are important for service management and optimization.

Authentication and Authorization:- Robust authentication and authorization helps to get effective Data Loss Prevention (DLP) solution. For every application, just user name and password is not most secure authentication mechanism. Sometime two factor or multi-factor authentication is needed [31]. We need to consider tiering access policies based on level of trust.

Infrastructure Hardening:- "Golden-image" VM and VM templates need to be hardened and cleaned [32]. This can be done while images are created. On

regular basis, testing of these master images need to be done.

End to end encryption:- IaaS as a service, both in public and private clouds, needs to take advantage of encryption from end-to-end. We can make use of whole disk encryption to encrypt all the data including user files on the disk. This prevents offline attacks. In addition to disk encryption, all communications to host OS and VMs in the IaaS infrastructure are encrypted. This can be done over SSL/TLS or IPSec.

2.3. Network issues on Cloud:

Network components are shared by different ternants due to resource pooling. Sharing resources allows attackers to launch across-tenant attacks [33]. The virtual networks increase the VMs interconnectivity, an important security challenge in cloud computing. The network is used to upload all the information. With the same aspect, [34] have stated security issues with network on cloud as a prime focus. It provides virtual resources, high bandwidth and software to the consumers on demand. But in reality, the network structure of this cloud faces various attacks and security issues like:

Browser Security:- Every client uses browser to send the information on network. The browser uses SSL technology to encrypt user's identity and credentials. But hackers from the intermediary host may acquire these credentials by the use of sniffing packages installed on the intermediary host. [35]. states that in order to overcome this, one should have a single identity but this credential must allow various levels of assurance which can be achieved by obtaining approvals digitally. Moreover, [36], has shown that Web Services security (WS-security) concept on browsers work with XML encrypted messages which does not need to be decrypted at intermediated hosts.

SQL Injection Attack- These attacks are malicious act on the cloud computing in which a spiteful code is inserted into a model SQL code. This allows the invader to gain unauthorized access to a database and eventually to other confidential information. Further, SQL injection attacks as described by [37] uses the special characters to return the data for example in SQL scripting the query usually ends up with where clause which again may be modified by adding more rows and information in it. The information entered by the hacker is misread by the website as that of the user's data and this will then allow the hacker to access the SQL server leading the invader to easily access and modify the functioning of a website. [38] have discussed in their paper on how the network related issues hinder the cloud computing and have also shown the SQL injection attack as the top intrusion detection.

Flooding Attacks- In this attack the invader sends the request for resources on the cloud rapidly so that the cloud gets flooded with the ample requests. As per the study carried out by [39]. flooding attack consume the critical system resources in order to paralyze the provided services and make them unavailable to its legitimate users in the cloud ecosystem.

XML Signature Element Wrapping:- It is found to be a very renowned web service attack. According to [40], it protects identity value and host name from illegal party but cannot protect the position in the documents. The attacker simply targets the host computer by sending the SOAP messages and putting any scrambled data which the user of the host computer cannot understand. As per the studies carried out by researchers at Ruhr University, and mentioned by the editor [41], the XML Signature wrapping attack changes simply the content of the signed part of a message without tampering the

signature. This would not let the user to understand the twisted data, thus misguiding and misleading the user.

Incomplete Data Deletion:- Incomplete data deletion is treated as hazardous one in cloud computing. According to [42], when data is deleted, it does not remove the replicated data placed on a dedicated backup server. The operating system of that server will not delete data unless it is specifically commanded by network service provider. Precise data deletion is majorly impossible because copies of data are saved in replica but are not available for usage.

Locks in- Another issue is locks in; at this time there is a small tender in the manner of tools, standard data format or procedures, services edge that could undertake data, application and service portability. This will not enable the customer to shift from one cloud provider to another or shift the services back to home IT location [43].

2.4. Virtualization: It has been the underlying concept towards such a huge rise of cloud computing in the modern era. It allows the cloud users to create, copy, share, migrate, and roll back virtual machines, which may allow them to run a variety of applications [44]. However, it also introduces new opportunities for attackers because of the extra layer that must be secured. Virtualized environments are vulnerable to all types of attacks for normal infrastructure and [45] stated that that security is agreater challenge as virtualization adds more points of entry and more interconnection complexity. The sub-section discusses the security weakness in cloud virtualization technology.

Securing VM images repository: Unlike physical servers, VMs are still under risk even when they are offline. VM images can be compromised by injecting malicious codes in the VM file or even stole the VM file itself. Secured VM images repository is the responsibilities of the cloud providers. Another issue related to VM templates is that such templates may retain the original owner information which may be used by a new consumer.

VM Hopping:- [46] and [47], stated that, with VM hopping, an attacker on one VM gains rights to use another victim VM. The attacker can check the victim VM's resource procedure, alter its configurations and can even delete stored data ,thus putting it in danger the VM's confidentiality, integrity and availability. With multi-tenancy, it makes the impact of a VM hopping attack larger than in a conventional IT environment. Because quite a few VMs can run at the same time and on the same host there is a possibility of all of them becoming a victim VMs. VM hopping is thus a critical vulnerability for IaaS and PaaS infrastructures.

Hypervisor security:- A hypervisor is responsible for virtual machines isolation; therefore , if the hypervisor is compromised, its virtual machines may potentially be compromised as well. The hypervisor is alow-level software that controls and monitors its virtual machines, so as any traditional software it entails security flaws [48]. Hypervisor security is the responsibility of cloud providers and the service provider. In this case, the SP is the company that delivers the hypervisor software such as VMware or Xen.

Side channel attacks:- An emerging concern for cloud delivery models using virtualization platforms is the risk of side channel attacks causing data leakage across co-resident virtual machine instances. However, it is possible that attackers who fail to compromise endpoints or penetrate cloud

infrastructure from outside the cloud perimeter, may consider this technique - acting as a rogue customer within a shared cloud infrastructure to access other customers' data.

CLOUD COMPUTING TAXONOMY CLASSIFICATION

The aim of this paper is to conduct holistic empirical study on cloud computing security in order to gain in-depth understanding of their security issues as far as the service is concerned. Figure 3, details the security issues reviewed in each and every model within the cloud computing ecosystem.

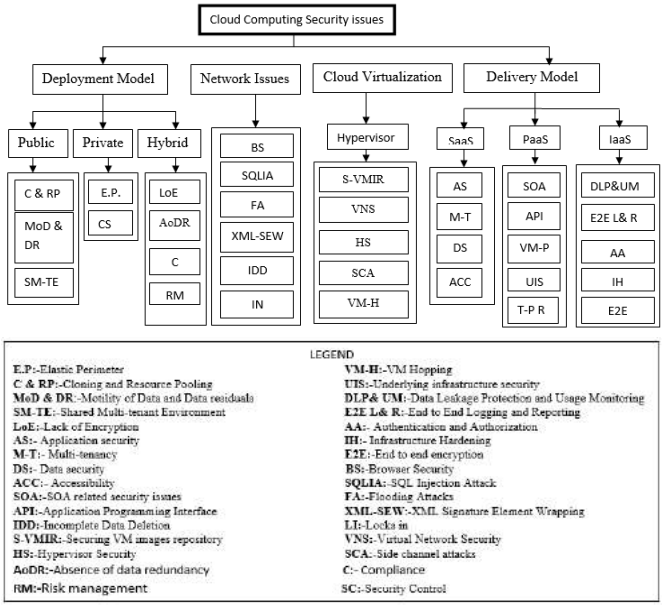


Figure 3. Security Taxonomy tree of cloud computing security issues

III. RELATED WORKS

Various security issues across cloud computing ecosystem as been reviewd by virous researchers on both the cloud provider and cloud user sides. The security responsivities of cloud providers include integrating solutions to ensure legitimate delivery of cloud services to cloud consumers. The security propositions that are necessary for the activities of cloud providers as reviewed in sub-section below.

Authentication and Authorization:- In his article [49] proposed a credential classification and framework for analyzing and developing solutions for credential management that include startegies to evaluate the complexity of cloud ecosystems. This study identifies a set of categories relevant for authentication and authorization for the cloud focusing on infrastructural organization and adapt these categories to the cloud context. In other work relevant with this security factor, a design model for multi-factor authentication in cloud computing environment is proposed [50] for analysis of the potential security threats. Another authentication solution is seen with MiLAMob [51], which provides a SaaS authentication middleware for mobile consumers of IaaS cloud applications. MiLAMob is a middleware-layer that handles the real-time authentication events on behalf of consumer devices with minimal HTTP traffic. FermiCloud [52] uses another approach for authentication and authorization - it utilizes public key infrastructure (PKI) X.509 certificates for user identification and authentication. Authorization-as-a-service (AaaS) approach using a formalized multi-tenancy authorization system, and providing administrative control over enhanced fine-grained trust models was introduced by [53]. They further propose use of cryptographic RBAC to enforce authorization policies regarding the trustworthiness of roles that are evaluated by the data owner.

Identity and Access Management:- The important functionalities of identity management systems for the success of clouds in relation to consumer satisfaction is discussed in [54]. He further presented an authorization system for cloud federation using Shibboleth - an open source implementation of the

security assertion markup language (SAML) for single sign-on with different cloud providers. This solution demonstrates how organizations can outsource authentication and authorization to thirdparty clouds using an identity management system. [55] also propose an integral federated identity management for cloud computing.

Confidentiality, Integrity, and Availability:- Both [56, 57] proposed a design that enables users to verify the integrity of VMs in the cloud. The proposed solution is called the trusted cloud computing platform (TCCP). In this approach, all nodes run a trusted virtual machine monitor to isolate and protect virtual machines. A TCCP guarantees confidentiality and integrity in data and computation and it also enables users to attest to the cloud service provider to ensure whether the services are secure prior to setting up their VMs. In 2011, [58] proposed CloudProof as a secure storage system to guarantee confidentiality, integrity and write-serializability using verifiable proofs of violation by external third parties. Confidentiality is ensured by private keys that are known only to the owner of the data that is to be encrypted. The main idea behind CloudProof is the use of the attestation mechanism. Attestations provide proof of sanity of users, data owners and cloud service providers. The attestation structure implements a solution called “block hash” for performing integrity checks through signature verification.

Fuzzy authorization (FA) for cloud storage [59] is another flexible and scalable approach to enable data to be shared securely among cloud participants. FA ensures confidentiality, integrity and secure access control by utilizing secret sharing schemes for users with smartphones who are using the cloud services. [60] improve cloud service resilience using a load-balancing mechanism called brownout. The idea behind this solution is to maximize the optional contents to provide a solution that is resilient to volatility in terms of flash crowds and capacity shortages (through loadbalancing over replicas) when compared to other approaches that are implemented using response-time or queue length. In another effort [61] the authors proposed a synchronization mechanism for cloud accounting systems that are distributed. The run time resource usage generated from different clusters is synchronized to maintain a single cloud-wide view of the data so that a single bill can be created. The authors also proposed a set of accounting system requirements and an evaluation method which verifies that the solution fulfills these requirements.

Security Monitoring and Incident Response:- According to [62] presents a centralized monitoring solution for cloud applications consisting of monitoring the server, monitors, agents, configuration files and notification components. Redundancy, automatic healing, and multi-level notifications are other benefits of the proposed solution which are designed to avoid the typical drawbacks of a centralized monitoring system, such as limited scalability, low performance and single point of failure. A scalable distributed monitoring system for clouds using a distributed management tree that covers all the protocol-specific parameters for data collection was presented by [63]. Hypervisor-based cloud intrusion detection systems are a new approach (compared to existing host-based and network-based intrusion detection systems) that is discussed in [64]. The idea is to use hypervisor capabilities to improve performance over data residing in a VM.

Security Policy Management:- In [65] the authors propose a generic security management framework allowing providers of cloud data management systems

to define and enforce complex security policies through a policy management module. The user activities are stored and monitored for each storage system, and are made available to the policy management module. Users' actions are evaluated by a trust management module based on their past activities and are grouped as "fair" or "malicious". An appropriate architecture for security management which satisfies the requirements of policy definitions (such as flexibility, expressiveness, extendibility and correctness) has been implemented. The authors evaluated the proposed system on a data management system that is built on data storage.

The policy management as a service (PMaaS) to provide users with a unified control point for managing access policies in order to control access to cloud resources independently of the physical location of cloud providers was introduced by [66]. PMaaS is designed specifically to solve the issue of having multiple access control authorization mechanisms employed by cloud service providers that restrict the flexibility of applying custom access control to a particular service. For this purpose, the PMaaS architecture includes a policy management service provider that is the entry point for cloud users to define and manage the policies. The cloud service provider imports the user-defined policies and acts a policy decision point to enforce the user policies.

IV. CRITIQUES OF EXISTING PROPOSITIONS TO CLOUD COMPUTING

The propositions solution for the cloud computing were noted to have a number of security setbacks that render them ineffective in cloud computing adoption. To start with, Authentication and Authorization, authentication which is the process that allows the user to provide proof of his identity [67]. It is often done through the login method, based on the using of a username and a password. This static mechanism leaves the system vulnerable to attacks, since hackers can use many techniques, such as sniffing and guessing, to steal user passwords [68]. On the part of Identity and Access Management, using the cloud services, users can ealiy access their personal information and make it available to various services across the Internet. [69] stated that, an identity management (IDM) mechanism can help authenticate users and services based on credentials and characteristics. A key issue concerning IDM in cloud is interoperability drawbacks that could result from using different identity tokens and identity negotiation protocols. Existing password-based authentication has an inherited limitation and poses significant risks. An IDM system should be able to protect private and sensitive information related to users and processes. However, multitenant cloud environments can affect the privacy of identity information and isn't yet well understood. Confidentiality, integrity and availability losses can make a big impact in the business of the cloud computing because the data is the core component for any business. Confidentiality is one of the prime constraints for the growth of cloud computing paradigm. Users when selecting the Cloud provider must be sure that the data that is given to the provider must be confidential. Provider must protect it from other users as well as must provide surety that even provider will also not peep into the data. Typically confidentiality is maintained by the encryption of the data that has been uploaded on the server of provider. But encryption has huge drawback in performance of the system. The integrity of data within complex cloud hosting environments such as SaaS configured to share computing resource amongst customers could provide a threat against data integrity if system resources are effectively segregated. On its part, security monitoring and incident response, many companies rely on third-party cloud

services provides and may not have access to every layer in the cloud computing stack, and therefore can't gain full visibility to monitor for potential security flaws and vulnerabilities. Lastly, on Security Policy Management, according to [70], stated that even though SaaS is based on autonomous agent, its security policy is still based on pre-defined rules, which limits the detection capabilities only to those attacks that are already known with the loud computing.

V. PROPOSED SECURITY MODEL OF CLOUD COMPUTING

The study proposed a cloud model mapped with security parameter used to enhanced the data and privacy of cloud computing. The layers in the proposed cloud security model includes; first the user creates a local agent, and establish a temporary security certificate then use this certificate for secure authentication in an effective period of time. The certificate includes the credentials such as the host name, user name, user id, start time, end time and security attributes etc. which are used access and authorization in cloud computing. Second, when the user's task use the resource on the cloud service layer, mutual authentication take place between user agent and specific application, while the application check if the user agent's certificate is expired, a local security policy is mapped. Third, according to user's requirements, cloud application will generate a list of service resource, and then pass it to the user agent. Through security API, user agent connects specific services. The cloud connection security ensures the safety of resources provided by the resource layer. The security API in this model should be achieved with SSL method, while the realization of cloud connection security uses SSL and VPN methods. The proposed security model is given in figure 4. The model consists the following security parameters:

Verification and Validation- This unit is required in cloud computing not only to authenticate users but also to ensure the accuracy of data and services on the cloud. The significance of security module is that cloud computing position is reachable by several customers and providers which want to use or provide many services and applications. Cloud service providers need to prove to the users that the services and data are valid, for example, appropriate signature algorithms. Consequently, user will be able to verify the authenticity of facts and services made available to them through digital signature. This protection part can also provide work for techniques such as One Time Password [71].

Security Policies:- Security policies are the basis of a resonance safety completion. Frequently organizations implement technical security solutions without creating foundation of policies, standards and security policies on firewall. Standards, procedures and guidelines referred to as policy in the superior sense of a worldwide information security policy [72]. Privilege Control- This security component is necessary to control cloud usage by different individuals and organizations. It protects user's privacy and ensures data integrity and secrecy by applying an anthology of rules and policies. Cloud users are granted different levels of access permissions and resource ownerships based on their account type. Only authorized users can access the authorized parts of the encrypted data through identity-based decryption algorithm. For example, in a healthcare cloud, not all practitioners have the same privileges to access patient's data, this may depend on the degree to which a practitioner is involved/specialized in treatment; patients can also allow or deny sharing their information with other healthcare practitioners or hospitals [73]. Encryption/Decryption algorithms [74] such as AES

and RC4 can be employed by this component to achieve confidentiality of information.

Data Protection- Data stored in the cloud storage resources may be very sensitive and critical, for example, clouds may host electronic healthcare records (EHR) which contain patients' private information and their health history [75]. They may also store critical banking information (e.g., clients account numbers, balances and transactions) or national security information. Cloud security model must protect data loss or injure by provide safe storage servers. These servers should also secure data retrieval and removal from the cloud. Securing data storage and processing is important since cloud users have no idea about data location. Techniques for data protection for example truncation, redaction, obfuscation, and others are able to be used in this security component. Encryption techniques can also be employed for data security. Hash functions and Message Authentication Code (MAC) can be employed in this unit to provide data integrity [76].

Security Services- The additional factors that directly affect cloud software assurance include authentication, authorization, auditing and accountability are used in cloud security services) [78]. Security-as-a- service is an industry form which a service contributor integrates security services into a commercial infrastructure on a subscription basis.

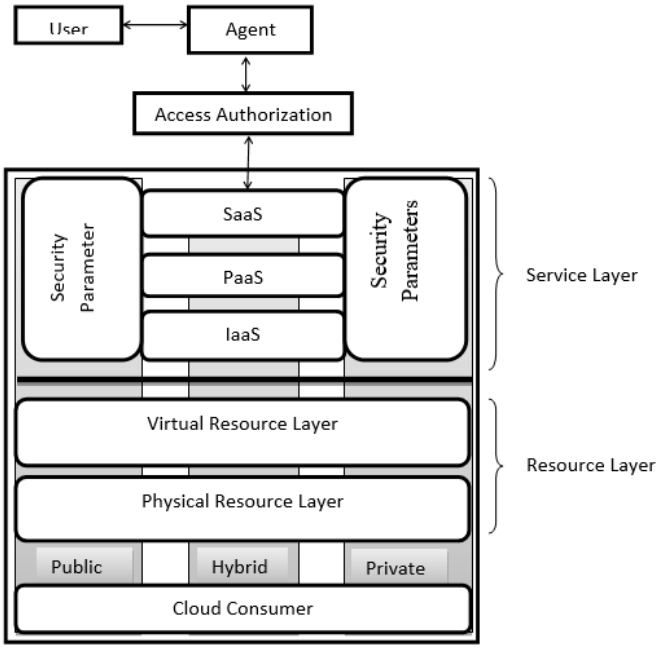


Figure 5. Proposed Security Model of Cloud Computing

Security-as a-service has applications such as anti-virus software delivered over the internet however the term can in addition pass on to security administration provided in-house by an external organization.

Threats / Attacks Detections- Clouds are vulnerable to many attacks and malicious behaviors that threaten both data and physical and virtual computing resources of the cloud. Basically, any set of actions that threaten the cloud security requirements (e.g., integrity, confidentiality and availability) are considered to be attacks. Attacks detection and prevention components are installed within the cloud security system to protect cloud resources from various anomalies. For example, denial-of-service attacks should be reduced to the minimum to guarantee the maximum availability of business, government, health and other critical information and services. This can be achieved by deploying technologies that provide high availability such as dynamic server load balancing and active/deactivate clustering [79]. Standard Distributed

Denial of Services (DDoS) mitigation techniques such as synchronous cookies and connection limiting can also be used. There are provisions for the next generation of intrusion detection systems and firewalls in order to protect the resources from intruders, viruses and malware [80].

VI. CONCLUSION

The aim of this paper is to carry out a holistic empirical study on cloud computing ecosystem security issues which hinders its adoption. It has been noted that most of the models such as delively model, service model and the enabling technologies of cloud computing that include virtualization and network are exposed to various security issues. The study further reviewed some of the security propositions of cloud cloud computing and their vulnerabilities. Based on these security setbacks, a novel security model has been proposed with some key parameters that can maintain both data integrity and confidentiality by deterring an illicit access to cloud data hence improve security.

VII. REFERENCES

[1]. P. Mell and T. Grance, (2011). The NIST Definition of Cloud Computing." National Institute of Standards and Technology, Sep-2011.

[2]. T. Rajesh and P. Vihari, (2013). Efficient Appraisal of Cloud Computing Through Comprehensive Confrontation of Security Issues and Discrepancies Involved. International Journal of Engineering Trends and Technology (IJETT) - Volume4Issue5- May 2013

[3]. O. Kuyoro, F. Ibikunle and O. Awodele (2011). Cloud Computing Security Issues and Challenges. International Journal of Computer Networks (IJCN), Volume (3) : Issue (5) : 2011

[4]. G. Shivani, (2017). Reviewing Security Concerns in Cloud Environment. International Journal of Computer Science and Mobile Computing, Vol.6 Issue.6, June- 2017, pg. 200-206

[5]. K. Wagh, (2014). Securing Data Transfer in Cloud Environment. Journal of Engineering Research and Applications www.ijera.com ISSN : 2248-9622, Vol. 4, Issue 5(Version 2), May 2014, pp.189-193

[6]. G. Bernd et al, (2011). Understanding Cloud Computing Vulnerabilities. IEEE, 1540- 7993/11, 2011, pp: 50-57

[7]. R. Balasubramanian and M. Aramudhan, (2012). Security Issues: Public vs Private vs Hybrid Cloud Computing. International Journal of Computer Applications (0975 – 8887) Volume 55– No.13, October 2012

[8]. D. Dikaiakos, et al. (2009). "Cloud computing: Distributed internet computing for IT and scientific research." Internet Computing, IEEE 13(5): 10-13.

[9]. Wayne A. Pauley,(2010) .Cloud Provider Transparency – An empirical evaluation. IEEE computer and reliability societies, IEEE, November 2010, pp: 32 – 39.

[10]. P. Wayne, (2010). Cloud Provider Transparency – An empirical evaluation the IEEE computer and reliability societies. IEEE, November 2010, pp: 32 – 39.

[11]. T. Hassan et al. (2010). Security and Privacy Challenges in Cloud Computing Environments. IEEE security and privacy, www.computer.org/security, 2010, pp. 24 – 31

[12]. R. Kui et al, (2010). Security Challenges for the Public Cloud. IEEE Press, 2012, pp. 69 – 73.

[13]. T. Hsin-Yi, (2012). Threat as a Service? Virtualization's impact on Cloud Security. IEEE, IT Pro, 2012, pp: 32- 37

[14]. R. Balasubramanian and M. Aramudhan (2012). Security Issues: Public vs Private vs Hybrid Cloud Computing. International Journal of Computer

Applications (0975 – 8887) Volume 55– No.13, October 2012.

[15].W. Cong et al, (2011). Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data. IEEE transactions on parallel and distributed systems, IEEE, Digital Object Indentifier 10.1109/TPDS.2011.282, 2011, pp: 1 – 14

[16].B. Marjory,(2010) .Hide and Seek in the Cloud. IEEE, March – April 2010, pp: 57-58.

[17].P. Mell, and T. Grance, (2009). The NIST definition of cloud computing, Recommendations of the National Institute of Standards and Technology Special Publication 800-145, National Institute of Standards and Technology, 2009.

[18].D. Owens, (2010). Securing elasticity in the Cloud. Commun ACM 53(6):46–51

[19].B. Tang, R. Sandhu, and Q. Li, (2013). Multi-tenancy authorization models for collaborative cloud services. in Collaboration Technologies and Systems (CTS), 2013 International Conference on, pp. 132–138, May 2013.

[20].S. Subashini and V. Kavitha (2011). A survey on Security issues in service delivery models of Cloud Computing. J Netw Comput Appl 34(1):1–11

[21].Ju et al, (2010) Research on Key Technology in SaaS. In: International Conference on Intelligent Computing and Cognitive Informatics (ICICCI), Hangzhou, China. IEEE Computer Society, Washington, DC, USA, pp 384–387.

[22].Grobauer, T. Walloschek, and E. Stocker E (2011). Understanding Cloud Computing vulnerabilities. IEEE Security Privacy 9(2):50–57

[23].Cloud Security Alliance (2012). Security guidance for critical areas of Mobile Computing. Available: https://downloads.cloudsecurityalliance.org/initiatives/ mobile/Mobile_Guidance_v1.pdf

[24].M. Leandro et al,(2012). Multitenancy authorization system with federated identity for cloud-based environments using shibboleth. in Proceedings of the 11th International Conference on Networks, ICN 2012, pp. 88–93, 2012.

[25].M. Stihler et al, (2012). Integral federated identity management for cloud computing. in New Technologies, Mobility and Security (NTMS), 2012 5th International Conference on, pp. 1–5, May 2012.

[26].Z. Wenjun,(2010). Integrated Security Framework for Secure Web Services. IITSI 2010, pp. 178-183

[27].B. Wang, et al (2009). Open Identity Management Framework for SaaS Ecosystem. ICEBE '09. pp. 512-517.

[28].B. Sumitra, C. Pethuru, and M. Misbahuddin, (2014). A Survey of Cloud Authentication Attacks and Solution Approaches. International Journal of Innovative Research in Computer and Communication Engineering, pp. 1-9, 2014.

[29].C. Klein, (2014). Improving cloud service re silience using brownout-aware loadbalancing. in Reliable Distributed Systems (SRDS), 2014 IEEE 33rd International Symposium on, pp. 31–40, Oct 2014.

[30].R. Popa, (2011). Enabling security in cloud storage slas with cloudproof. in Proceedings of the 2011 USENIX Conference on USENIX Annual Technical Conference, USENIX ATC'11, (Berkeley, CA, USA), pp. 31–31, USENIX Association, 2011.

[31].L. Almutair and H. Zaghloul, (2013). In Proceedings of the The Third International Conference on Digital Information Processing and Communications (ICDIPC '13), pp. 676–686, UAE, 2013.

[32].J. Shropshire, (2014) "Analysis of monolithic and microkernel architectures: Towards secure hypervisor design," in Proceedings of the 47th

Hawaii International Conference on System Sciences, HICSS 2014, pp. 5008–5017, usa, January 2014.

[33]. N. Santos, K. Gummadi, and R. Rodrigues, (2009). Towards trusted cloud computing. in Proceedings of the 2009 Conference on Hot Topics in Cloud Computing, HotCloud'09, (Berkeley, CA, USA), USENIX Association, 2009.

[34]. R. Banyal, P. Jain, and V. Jain,(2013). Multi-factor authentication framework for cloud computing. in Computational Intelligence, Modelling and Simulation (CIMSim), 2013 Fifth International Conference on, pp. 105–110, Sept 2013.

[35]. P. Tabakki et al. (2010). SecureCloud: Towards a Comprehensive Security Framework for Cloud Computing Environments. Proc. 1st IEEE Int'l Workshop Emerging (2012

[36]. K. Steve et al, (2012). The Future of Authentication. IEEE, January-February 2012, pp: 22 – 27

[37]. M. Jensen, (2009). On Technical Security Issues in Cloud Computing. IEEE International Conference on Cloud Computing, pp: 109 – 116.

[38]. Q. Sara and K. Kausar, (2012)"Cloud Computing: Network/Security Threats and counter measures", Interdisciplinary Journal of Contemporary Research in Business, ijcrb.webs.com, January 2012, Vol 3, N0 9, pp: 1323 – 1329.

[39]. S. Roschke, et al., (2009). Intrusion Detection in the Cloud. IEEE International Conference on Dependable, Autonomic and Secure Computing, Chengdu, China, 2009.

[40]. A. Elmustafa and E. Rasha, (2015). Network Denial of Service Threat Security on Cloud Computing A Survey. IJSRSET | Volume 1 | Issue 5 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099

[41]. D. Perez-Botero, J. Szefer, and R. Lee, (2013). Characterizing hypervisor vulnerabilities in cloud computing servers," in Proceedings of the 2013 International Workshop on Security in Cloud Computing, Cloud Computing '13, (New York, NY, USA), pp. 3–10, ACM, 2013.

[42]. L. Ertaul, S. Singhal and S. Gökay, (2010). Security challenges in Cloud Computing. In: Proceedings of the 2010 International conference on Security and Management SAM'10. CSREA Press, Las Vegas, US, pp 36–42

[43]. Q. Sara and K. Kausar, (2012)"Cloud Computing: Network/Security Threats and counter measures", Interdisciplinary Journal of Contemporary Research in Business, ijcrb.webs.com, January 2012, Vol 3, N0 9, pp: 1323 – 1329.

[44]. G. Grispos, B. Glisson, and T. Storer, (2013). Cloud Security Challenges: Investigating Policies. Standards, and Guidelines in a Fortune 500 Organization", 21st European Conference on Information Systems, 5-8, 2013

[45]. A. Jasti et al, (2010) Security in multi-tenancy cloud. In: IEEE International Carnahan Conference on Security Technology (ICCST), KS, USA. IEEE Computer Society, Washington, DC, USA, pp 35–41.

[46]. Reuben JS (2007) A survey on virtual machine Security. Seminar on Network Security. http://www.tml.tkk.fi/Publications/C/25/papers/Reuben_final.pdf. Technical report, Helsinki University of Technology, October 2007.

[47]. K. Owens, (2009). Securing Virtual Computer Infrastructure in the Cloud. white paper, Savvis Communications Corp., 2009

[48]. A. Jasti et al.,(2010) "Security in Multi-Tenancy Cloud," Proc. IEEE Int'l Carnahan Conf. Security Technology (ICCST 10), IEEE Press, 2010, pp. 35–41.

[49].H. Takabi and J. Joshi, (2012). Policy management as a service: An approach to manage policy heterogeneity in cloud computing environment," in System Science (HICSS), 2012 45th Hawaii International Conference on, pp. 5500–5508, Jan 2012.

[50].N. Mimura et al,(2013). A framework for authentication and authorization credentialsin cloud computing," in Trust, Security and Privacy in Computing and Communications (TrustCom), 2013 12th IEEE International Conference on, pp. 509–516, July 2013.

[51].R. Banyal, P. Jain, and V. Jain,(2013). Multi-factor authentication framework for cloud computing. in Computational Intelligence, Modelling and Simulation (CIMSim), 2013 Fifth International Conference on, pp. 105–110, Sept 2013.

[52].R. Lomotey and R. Deters,(2013). Saas authentication middleware for mobile consumers of iaas cloud. in Services (SERVICES), 2013 IEEE Ninth World Congress on, pp. 448–455, June 2013.

[53].H. Kim and S. Timm, (2014). X.509 authentication and authorization in fermi cloud. in Utility and Cloud Computing (UCC), 2014 IEEE/ACM 7th International Conference on, pp. 732–737, Dec 2014.

[54].B. Tang, R. Sandhu, and Q. Li, (2013). Multi-tenancy authorization models for collaborative cloud services. in Collaboration Technologies and Systems (CTS), 2013 International Conference on, pp. 132–138, May 2013.

[55].M. Leandro et al,(2012). Multitenancy authorization system with federated identity for cloud-based environments using shibboleth. in Proceedings of the 11th International Conference on Networks, ICN 2012, pp. 88–93, 2012.

[56].M. Stihler et al, (2012). Integral federated identity management for cloud computing. in New Technologies, Mobility and Security (NTMS), 2012 5th International Conference on, pp. 1–5, May 2012.

[57].N. Santos, K. Gummadi, and R. Rodrigues, (2009). Towards trusted cloud computing. in Proceedings of the 2009 Conference on Hot Topics in Cloud Computing, HotCloud'09, (Berkeley, CA, USA), USENIX Association, 2009.

[58].T. Garfinkel, (2003). A virtual machine-based platform for trusted computing. in Proceedings of the Nineteenth ACM Symposium on Operating Systems Principles, SOSP '03, (New York, NY, USA), pp. 193–206, ACM, 2003.

[59].R. Popa, (2011). Enabling security in cloud storage slas with cloudproof. in Proceedings of the 2011 USENIX Conference on USENIX Annual Technical Conference, USENIX ATC'11, (Berkeley, CA, USA), pp. 31–31, USENIX Association, 2011.

[60].S. Zhu and G. Gong, (2014). Fuzzy authorization for cloud storage," Cloud Computing, IEEE Transactions on, vol. 2, pp. 422–435, Oct 2014.

[61].C. Klein, (2014). Improving cloud service re silience using brownout-aware loadbalancing. in Reliable Distributed Systems (SRDS), 2014 IEEE 33rd International Symposium on, pp. 31–40, Oct 2014.

[62].E. Lakew, (2014). A synchronization mechanism for cloud accounting systems," in Cloud and Autonomic Computing (ICCAC), 2014 International Conference on, pp. 111–120, Sept 2014.

[63].M. Anand, (2012) .Cloud monitor: Monitoring applications in cloud. in Cloud Computing in Emerging

[64].Markets (CCEM), 2012 IEEE International Conference on, pp. 1–4, Oct 2012.

[65].A. Brinkmann, (2013). Scalable monitoring system for clouds. in Proceedings of the 2013 IEEE/ACM 6th International Conference on

Utility and Cloud Computing, UCC '13, (Washington, DC, USA), pp. 351–356, IEEE Computer Society, 2013.

[66].C. Basescu, (2011). Managing data access on clouds: A generic framework for enforcing security policies. in Advanced Information Networking and Applications (AINA), 2011 IEEE International Conference on, pp. 459–466, March 2011.

[67].H. Takabi and J. Joshi, (2012). Policy management as a service: An approach to manage policy heterogeneity in cloud computing environment," in System Science (HICSS), 2012 45th Hawaii International Conference on, pp. 5500–5508, Jan 2012.

[68].H. Chang and E. Choi, (2011). User authentication in cloud computing. in International Conference on Ubiquitous Computing and Multimedia Applications, pp. 338–342, 2011.

[69].B. Sumitra, C. Pethuru, and M. Misbahuddin, (2014). A Survey of Cloud Authentication Attacks and Solution Approaches. International Journal of Innovative Research in Computer and Communication Engineering, pp. 1-9, 2014.

[70].S. Jaydip (2013). Security and Security andPrivacy Privacy Privacy Issues in Cloud Computing Computing. Innovation Labs, Tata Consultancy Services Ltd., Kolkata, INDIA

[71].K. Issa , K. Abdallah and A. Muhammad (2014).Cloud Computing Security: A Survey. Cloud Computing Security: A Survey.

[72].E. Ahmed et.al. (2012). A Framework for Secure Cloud Computing. IJCSI International Journal of Computer Science, Vol. 9, issue 4, No 3, ISSN (Online): 1694-0814, July 2012.

[73].R. Krutz et.al. (2013) .CLOUD SECURITY Comprehensive Guide to Secure Cloud Computing. Chapter 2: Cloud Computing Architecture, WILEY-INDIA, ISBN : 978-81-265-2809-7.

[74].R. Zhang and L. Liu, (2010) .Security Models and Requirements for Healthcare Application Clouds. IEEE 3rd International Conference on Cloud Computing.

[75].Leonard D.C. et.al.(2009) "Realization of Universal Patient Identifier for Electronic Records through Biometric Technology", IEEE Trans on Information Technology in Biomedicine, Vol. 13, No. 14.

[76].E. Ahmed et.al.(2012). A Framework for Secure Cloud Computing. IJCSI International Journal of Computer Science, Vol. 9, issue 4, No 3, ISSN (Online): 1694-0814, July 2012.

[77].A. Tripathi and A.Mishra (2011). Cloud Computing Research Challenges. IEEE 5th International conference on Biomedical Engineering and Informatics, pp- 1397-1401.

[78].R. Krutz et.al.(2013). CLOUD SECURITY Comprehensive Guide to Secure Cloud Computing. Chapter 2 : Cloud Computing Architecture, WILEY-INDIA, ISBN : 978-81-265-2809-7.

[79].A. Tripathi and A. Mishra (2011). Cloud computing security considerations. IEEE International conference on signal processing, communication and computing (ICSPCC).

[80].A. Mohiuddin et.al.(2012). An Advanced Survey on Cloud Computing and State-of-the-art Research Issues. IJCSI International Journal of Computer Science Issues, ISSN (Online): 1694-0814, vol. 9, issue 1, No 1.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Suitable Solid Waste Disposal Site Selection Using Geographical Information System: A Case of Debre Markos Town, Ethiopia

Abinet Addis Mihiretie
Department of Civil & Environmental Engineering, Debre Markos University, Debre Markos, Ethiopia

ABSTRACT

In this study an effort has been made to identify suitable sites for safe disposal. For proper identification and selection of appropriate sites for solid waste disposal careful and systematic procedures need to be adopted and followed. The main objective of this research was identified the suitable solid waste disposal site by using the GIS-based approaches in Debre markos town. The present study had considered various factors like road networks; rivers, soil, slope, altitude and land use/ land cover for selecting a suitable solid waste disposal site within the study area. The relative weights of the factor were estimated using AHP and factor maps were developed by using GIS spatial operations. The final weighted factor map produced an overall solid waste disposal suitability map. The solid waste disposal site suitability map was presented in four suitability index such as highly suitable, moderately suitable, low suitable and unsuitable. The result shows that around 21% area is highly suitable for solid waste disposal site, 25% is moderately suitable, 28% of study area has low suitable and 26% area is unsuitable for solid waste disposal site.

Keywords : Debre markos, GIS, solid waste disposal site

I. INTRODUCTION

1.1. Back ground of the Study

Solid waste disposal is becoming a major global problem. Due to increasing human activity, solid waste is creating serious damage to the ecosystem and human health. Damage is caused by illegal dumping of urban waste in unacceptable locations (Ahmad et al., 2011). Uncontrolled growth of the urban population in developing countries in recent years has made solid waste management an important issue. Today municipal solid waste management is one of the most challenging issues in urban cities, which are facing a serious pollution problem due to the generation of huge quantities of solid waste (Debishree and Samadder, 2014).

Solid waste in Ethiopia the booming growth of cities of the developing world has outpaced the financial and manpower resources of municipalities to deal with provision and management of services, of which solid waste is the major one. Lack of these services greatly affects the urban poor, women and children who are vulnerable to health hazards. Twenty two human diseases are related to improper solid waste management (World Bank, 1999). Moreover, its effects are also reflected in reduced productivity, low income and poor quality of life and deteriorated environment. Similar to cities of most developing countries, provision of required services lags behind

the need and development of settlements in urban areas of Ethiopia. Integrated infrastructure and housing development is not widely practiced. Provision of solid and liquid waste collection and disposal is low (most urban areas lack the service). In addition to this, deterioration of the immediate environment in the households and their surrounding is increasing. With the current growth rate of urban population in Ethiopia, it is estimated that the population of most urban areas especially small urban centers is doubling every 15-25 years (Bedasa and Wondwossen, 2019). As solid waste generation increases with economic development and population growth, the amount in these urban areas will double within a similar time range. Municipalities in Ethiopia have to be prepared for this challenge (Ambaneh, 2016). An integrated urban rural development study undertaken in 1988 showed that among the eleven project towns: Addis Ababa, Akaki, Assela, Ambo, Arsi Negele, Goba, Mizan Teferri, Robe, Wolisso, Ziway, and Shashemene, only Addis Ababa had centralized waste disposal system (NUPI, 1989). The towns had no waste collection trucks; four of the municipalities assigned other vehicles to collect waste once or twice a week. Among those who have the service the coverage is very low, usually being limited to street and market cleaning. Recently, most municipalities in Ethiopia have become aware of the negative consequences of poor sanitation.

Geographical Information Systems (GIS) which are of the ability of processing and analyzing each kind of data, are of an important role for proper solid waste areas; in a word a decision support tool for determining alternative areas. Such as GIS makes it possible to include all kind of spatial technical, economic and social conditions and studying the results developed under different conditions. But another important question in solid waste site selection process is to find out the more proper alternatives among the proper solid waste (Iscan and Yagci, 2015).

In Debre markos town the source of solid wastes from hospitals, constructions, households and commercial areas are collected and dumped, which may contain leachable and toxic compounds that are harmful to the environment and human health. The disposal system of solid waste in the town is open dumping, which is commonly carried out on areas which are not been selected following any scientific way for suitability. Majority of population in Debre markos disposed waste illegally into public places, river and ditches.

In this study, the main objective was identified a suitable solid waste disposal site in Debre markos town using GIS techniques. Specifically identifying factors for selecting suitable solid waste disposal site; determining the relative importance of the factors for solid waste disposal site using a weighted linear combination (WLC) methods and mapping of suitable solid waste disposal site in the study area.

II. MATERIALS AND METHODS

2.1. Study Area

Debre markos town is located in the northwestern part of the country at a distance of approximately 299 kms by road from Addis Ababa on the Bahir Dar highway. The town is the center of the East Gojjam zone. According from UTM coordinate system, the location of the town is approximately between 356564m– 365423m East direction and 1136757m – 11470559m North direction as shown in Fig.1.

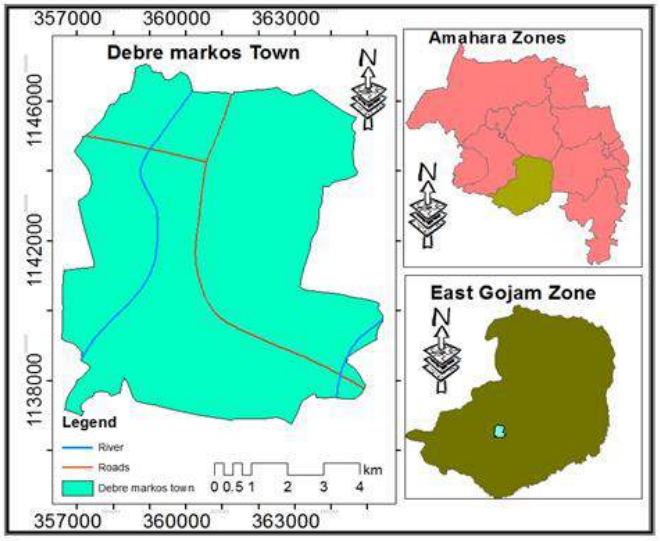


Fig. 1: Location map of the study area

2.2. Methods

Debre markos town was selected to identify solid waste disposal areas. For this Multi Criteria Analysis was used for creating various layers to be used in GIS domain to produce a single output map. The weights were developed by providing a series of pair wise comparisons of relative importance. Based on experience and likely impact on surrounding environment different weights were assigned to all the parameters. Weighted linear combination was used to produce the suitability of solid waste disposal map. As for the final weighted factor map is a weighted linear combination of factor maps, an equation (1) as following:

$$S = \Sigma w_i x_i \text{ where, } S = \text{suitability, } w_i = \text{weight of factor i and } x_i = \text{factor map i.} \quad (1)$$

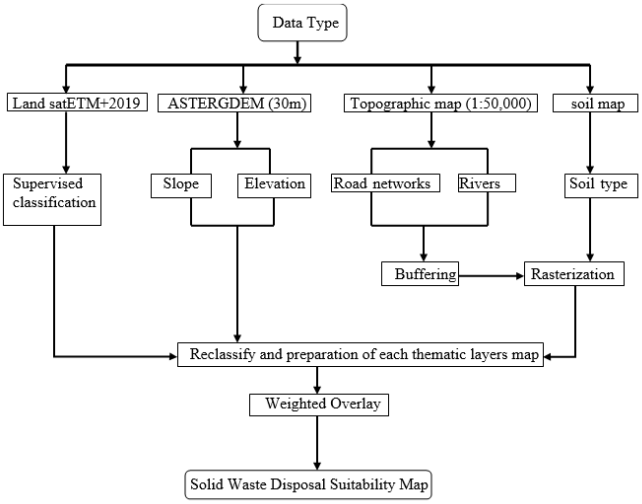


Fig. 2: Flow chart of the methodology

2.3. Data Type and Source of Data

For this study different types of datasets were integrated within GIS environment to come up with final output of solid waste site selection in Debre markos Town. The datasets used are tabulated below in **(table 1).**

Table 1: Data type and their sources

No.	Types of Datasets	Format	Sources
1	Land use land cover	Raster	Interpretation of Landsat 8 image (2019) from USGS website
2	Slope and Elevation	Raster	Interpretation of ASTER GDEM image (2019)
3	Soil types	Vector, Shape file	Minister of Water and Energy, Addis Ababa
4	Surface water	Vector, Shape file	Minister of Water and Energy, Addis Ababa
5	Road network	Vector, Shape file	Interpretation the study area of Topographical map from (EMA), Addis Ababa

© 2020 IJSRST | Print ISSN: 2395-6011 | Online ISSN: 2395-602X | Published : 18/12/2020-19-26

III. RESULTS AND DISCUSSION

No.	Types of Datasets	Format	Sources
1	Land use land cover	Raster	Interpretation of Landsat 8 image (2019) from USGS website
2	Slope and Elevation	Raster	Interpretation of ASTER GDEM image (2019)
3	Soil types	Vector, Shape file	Minister of Water and Energy, Addis Ababa
4	Surface water	Vector, Shape file	Minister of Water and Energy, Addis Ababa
5	Road network	Vector, Shape file	Interpretation the study area of Topographical map from (EMA), Addis Ababa

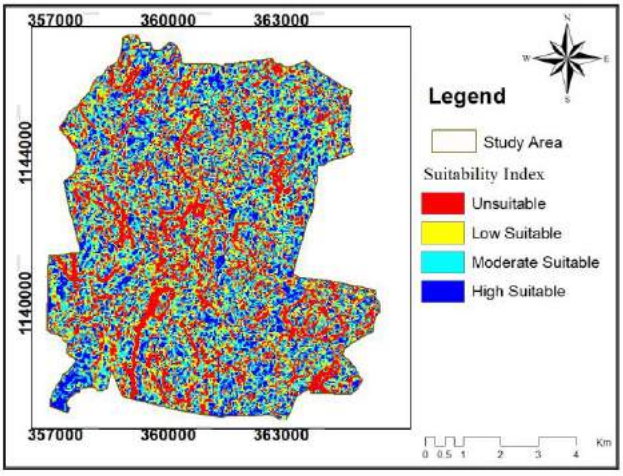


Fig. 3: Slope factor map

Table 2: Classification, ranking and weighting of Slope factor

Factor	Weigh t	Class	Rank	Degree of disposal
Slope (Degree)	0.10	0-10	4	High Suitable
		10-15	3	Moderate Suitable
		15-20	2	Low Suitable
		>20	1	Unsuitable

3.1.2. Altitude

Altitude was generated from the ASTER GDEM data. The altitude of the study area generally ranges between 2240m and 2514m. Altitude is a basic parameter to build a landfill. Areas with too high altitude are not suitable for solid waste disposal site because of it is difficult to permit gravity flow (Kao and Lin, 2005). The ideal places for landfill should be located at a low altitude in order to permit gravity flow.

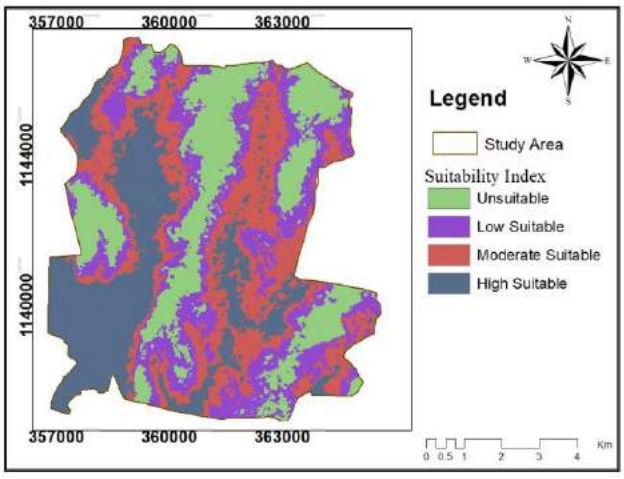


Fig. 4: Altitude factor map

Table 3: Classification, ranking and weighting of Altitude factor

Factor	Weight	Class	Rank	Degree of disposal
Altitude (meter)	0.10	2240-2383	4	High Suitable
		2383-2411	3	Moderate Suitable
		2411-2436	2	Low Suitable
		2436-2514	1	Unsuitable

3.1.3. Distance from Rivers

The solid waste disposal site should not be placed within surface water to protect surface water from contamination by leachate (Abdelhakim et al, 2016). Amhara National Regional State has established the Regional Hygiene and Environmental Executive Committee which has issued regional environmental sanitation regulations (No.16/2000); it clearly states that the solid waste disposal should not be dump to any water surface like river, stream, Lake etc. In the study area, Chemoga River is one of the main surface water. Researchers have suggested a distance up to 500m away from surface wate (Kontos et al., 2005).

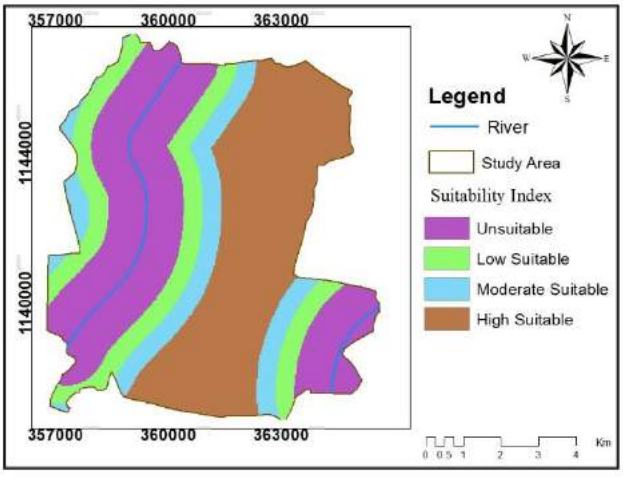


Fig. 5: Rivers factor map

Table 4: Classification, ranking and weighting of Altitude factor

Factor	Weight	Class	Rank	Degree of disposal
River (meter)	0.15	0-1000	1	Unsuitable
		1000-1500	2	Low Suitable
		1500-2000	3	Moderate Suitable
		>2000	4	High Suitable

3.1.4. Distance from Road Networks

The road is one of the important factors in considering a solid waste disposal site selection. There is no specific rule of what should be the best distance to place the solid waste disposal site. Most studies suggested that the disposal site should be located within a 500m buffer from the existing roads (Shahmoradi and Isalou, 2013). However, planners may prefer to give an aesthetic concern when deciding a location of solid disposal site. Also, the site should not be close to the existed road networks that affects the public health.

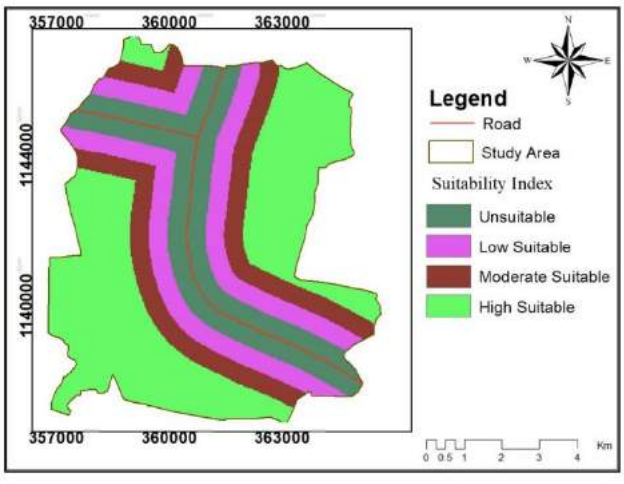


Fig. 6: Road networks factor map

Table 5: Classification, ranking and weighting of Road networks factor

Factor	Weight	Class	Rank	Degree of disposal
Road Networks (meter)	0.20	0-500	1	Unsuitable
		500-1000	2	Low Suitable
		1000-1500	3	Moderate Suitable
		>1500	4	High Suitable

Table 6: Classification, ranking and weighting of Land use Land cover factor

Factor	Weight	Class	Rank	Degree of disposal
Land use Land cover	0.35	Water body and forest	1	Unsuitable
		Built up and vegetation area	2	Low Suitable
		Agricultural area	3	Moderate Suitable

3.1.4. Land use Land cover

In the present study land use/land cover was analyzed from Landsat 8 image data for the solid waste disposal site selection. In the study area, there are five major land uses such as water body, built-up area, forest and agricultural land and vegetation areas. By reviewing different literature, it was advisable to select land, it was occupied by bare and grass lands a high suitable area for solid waste disposal.

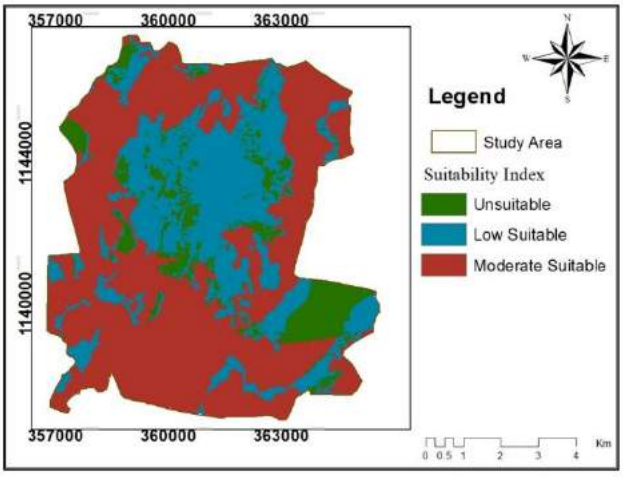


Fig. 7: Land use Land cover factor map

3.1.5. Soil Types

There were three type of soils texture in the study area based on FAO (2006) classification such as eutric nitisols, Pellic vertisols and chromic vertisols. Soil should be of sufficiently low permeability to significantly slow the passage of leachate from the site. The higher the permeability of the textured unit, the lower is the site suitability for landfill practices and the higher is its vulnerability. In the present study, soil texture was classified into three zones according to the texture unit of permeability characteristics.

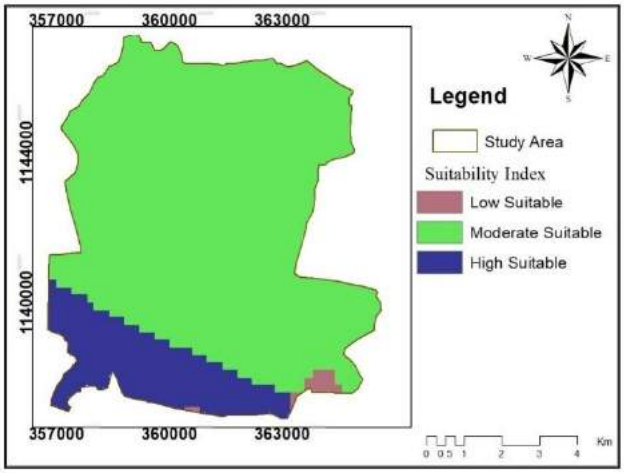


Fig. 8: Soil types factor map

Table 7: Classification, ranking and weighting of Soil types factor

Factor	Weight	Class	Rank	Degree of disposal
Soil types	0.10	eutric nitisols	2	Low Suitable
		Pellic vertisols,	3	Moderate Suitable
		Chromic vertisols	4	High Suitable

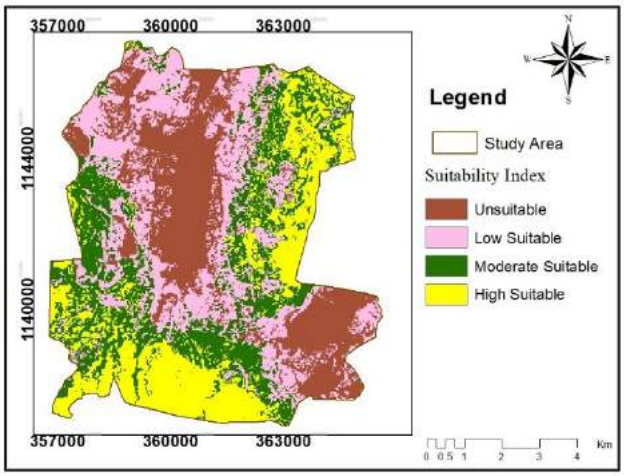


Fig. 9: Solid waste disposal suitability Map of Debre markos town

3.1. Overlay Analysis and Identifying Suitable Areas

To determine the suitable areas for solid waste disposal sites, there are six factors were considered. The factors are surface water; road networks, soil texture, land use land cover, slope and altitude. Each factor thematic map was prepared using ArcGIS spatial analyst tools, and prepared maps were converted into raster format. The overlay analysis of the final weighted factor map was produced the final suitable solid waste disposal site map was prepared using ArcGIS software by the following principle:

Suitability Map = Σ [(Factor map * weight) (2)

Suitability index = ([surface water map] * 0.15 + [land use map]*0.35+ [slope map]*0.10+ [altitude map]*0.10 + [soil map]*0.10 + [road map]*0.20)

The result of weighted factor map in the ArcGIS software reveals four suitability indices for the municipal solid waste dumping site of Debre markos town. These are unsuitable, low, moderate and high suitable.

The area coverage of each suitability index of the sites was calculated in ArcGIS environment showed that (26%) of the study area is unsuitable for solid waste disposal site as the areas are environmentally unfavorable and economically impracticable to be proposed as solid waste disposal site. This unsuitable area include close to surface water (area with a 500m buffer zone), areas with steep slope (>20^0), a higher altitude (>2514m), water bodies and forest areas and close to road networks (with a 500m buffer zone), respectively. The main advantage of these areas unsuitable was to minimize their negative effects of on environment and public health as well as to minimize the cost of construction and maintenance of the land fill. However, (27%) of the area was low suitable and the area of (25%) moderate suitable. Out of the remaining areas of (21%) are high suitable, these areas are preferable for solid waste disposal sites, because of their minimum effect on environment, public health and cost effective than other parts of the study area.

Table 8 : Statistical analysis for the solid waste disposal site suitability

Suitability Index	Area in (km^2)	Area in (%)
Unsuitable	233.51	26
Low Suitable	246.09	28
Moderate Suitable	228.04	25
High Suitable	192.35	21
Total	**899.99**	**100.00**

IV. CONCULUSION

The analysis factors were used in the identification of suitable sites for solid waste disposal in Debre markos town using the GIS techniques. Currently solid waste is disposed at open spaces near to outside the municipal boundary impact on health hazards for the rural settlements. Geographic information system tools widely used to assist the identification of suitable sites for solid waste disposal. In this study, Selection of suitable site depends on a number of factors were considered. The factors are slope, altitude, road networks, rivers, soil types and land use land cover but some other factors like ground water, settlements, and geology that may also affect the sitting of suitable sites but they were not included in the present study due to data limitations. After produced the final weighted map, the high suitable solid waste disposal sites were identified in the study area. The high suitable solid waste disposal sites were located in the southern and eastern part of a town.

V. REFERENCES

[1]. Shahmoradi B.and Isalou A. (2013). Site selection for wastewater treatment plant using Integrated fuzzy logic and multicriteria decision model: A case study in Kahak, Iran. Adv Environ Health Res; 1(1): 51-61.

[2]. Abdelhakim El Maguiri, B. K. (2016). Landfill site selection using GIS, remote sensing and multicriteria decision analysis: case of the city of Mohammedia, Morocco. DOI 10.1007/s10064-016-0889.

[3]. Ahmad A., B. A. (2011). GIS-Based Weighted Linear Combination Analysis and Remote Sensing Techniques to Select Optimum Solid Waste Disposal Sites within Mafraq City, Jordan. Journal of Geographic Information System; 19(3): 267-278.

[4]. Wondwosen M. and Bedasa A. (2019). Suitable Solid Waste Disposal Site Selection Using Geographical Information System (GIS): A Case of Debre Berhan Town, Ethiopia. 7(1).American Journal of Environmental Protection, Vol. 7, No. 1, 17-23 .

[5]. Kao J. and Lin H.Y. (2005). Grid-based heuristic method for multifactor landfill siting. Journal of Computing in Civil Engineering; 19(4): 369-376.

[6]. Kontos T. D., K. D. (2005). MSW landfills with a spatial multiple criteria analysis methodology, Waste Manag; 25:818–832.

[7]. Ambaneh M. (2016). Solid Waste Disposal Site Selection Using Gis And Remote Sensing: For Mojo Town, Ethiopia.

[8]. Samadder, D. K. (2014). Application of GIS in Landfill Siting for Municipal Municipal Solid Waste. 4(2249-3131 Volume 4, Number 1, pp. 37-40).

[9]. Iscan and Yagci (2015). Selection of Solid Waste Landfill Site Using Geographical Information System (GIS). 9(8).

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Design and Implementation of a Rule Based Afaan Oromoo Conversational Chatbots

Fikadu Wayesa

Wachemo University, School of Computing, Information Technology Department, Ethiopia

ABSTRACT

Chatbot technologies are very popular one in the field of NLP and AI to process basic expressions of interaction between humans (end users) and devices as Question Answering System (QAS). They are used in many ranges like in industry to provide customer services based on their need to reach on their desires, transportation, health, entertainment, education, psychology to monitor human emotions and in many. Messaging service is a popular means of communication as text based messaging (SMS) or through messaging applications. Chatbots are virtual assistant technologies that are designed and developed based on the user's query and to their language. The aim of this paper is to develop and implement a rule based conversational chatbot for Afaan Oromo Language. The approach gets a simple Afaan Oromoo query as input and retrieves answers from a set of predefined rules or responses. The performance of our model is evaluated through human judgment.

Keywords: Afaan Oromoo, Chatbot, NLP, Conversation, Rule-Base

I. INTRODUCTION

In the digital era abundant amount of end user services interactions are helping us to solve our daily activities. In the field of Artificial Intelligence (AI), Chatbots are technologies designed for making conversations to deliver answers quickly to user's need in short amount of time as human-human interaction to assist them to accomplish a task (Martin 2018).

These systems are used in everywhere to engage customer service such as deliver flight information,

customer support or in psychological counseling. It assists the user by asking series questions and guide them to what the user is looking for.

Basically, there are three forms of Chatbots: Rule-based (fixed rule) chatbot, Self-Learning and Generative-based chatbots. Self-learning is also called AI-Based Chatbots (Martin 2018). We can combine these forms of chatbots and we can implement an approach called Hybrid.

The rule-based methods depend on a list of simple predefined queries and possible resultant answers. It does not need any machine learning approach and language processing is not mandatory. They are intended for simple queries and they may fail for more complicated questions since they can't produce their own responses. This form of chatbots has one-to-one input and replies. So a bot would be trained according to the defined rules (Soufyane Ayanouz April 2020).

Self-Learning chatbots depend on machine learning and word vectorization. Words are vectorized, because machine learning methods use numbers for prediction. These bots follow the machine learning approach which is rather more efficient and is further divided into two more categories Retrieval-Based Models (corpus-based) and Generative Models (O'Reilly 2017).

Retrieval-based chatbot uses empirical to select a response from predefined responses. It can handle context of the query and selects the best answer from predefined answer.

Generative-based chatbots are the most complex compared to the two previous approaches. It also called cognitive based chatbots. It requires a much bigger data set and use deep learning methods for training. They have cognitive ability to generate different answers (O'Reilly 2017)

II. RELATED WORKS

ELIZA chatbot dialog system was developed to stimulate a Rogerian psychologist, based on a branch of clinical psychology whose methods involve drawing the patient out by reflecting patient's statements back at them based on the predefined rule (Weizenbaum 1996).

(D'ÁVILA 2018) Designates a rule based chatbot development for long conversations using KINO dataset. The goal of the paper was to match between a response candidates in a given conversational speeches. They proposed a framework in which the possible to verify that the positively associated words were part of the information in which the bot had been programmed to search in its knowledge base. A manual classification was necessary for the words negatively related to understand and identify the types of issues in the identification of the user's intentions.

(Jagdish Singh 2019) Proposed a Rule-based chabot for Asian Pacific University (APU) students to answer their enquiries from their administrative offices. The designed chatbot permits end users to observe information and interact with the chatbot regarding their enquiries. The developed system was compatible with bot's app page on Facebook for access via laptop or computer devices.

There is no work attempted work for Afaan Oromo Language in the area of chatbot system. This language is rich in its structure, grammar and semantics.

Here, this research work focused on the following questions.

1. What are the possible methodologies used in the ChatBot implementation system?
2. What challenges are there in the developing rule based chatbot?

The Model Architecture

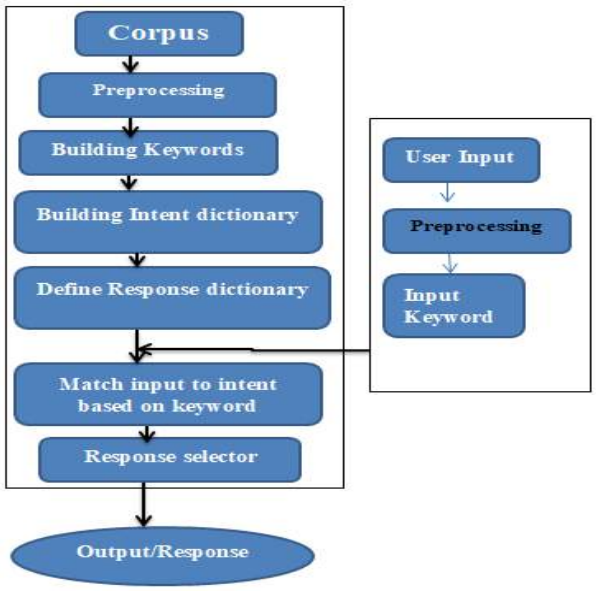


Figure 1 Proposed System Architecture

III. Experimental Results

A. Components of the Model

Corpus

We have collected a lists of query with their possible response that are used in Afaan Oromo Conversational System.

Preprocessing

Data should be preprocessed first in order to design any application, the first thing we need is data, to attain a particular format to make machine understandable. To get the possible answer for the query a number of NLP tasks have been applied to our corpus and to the query. Generally, the following tasks have been applied.

1. The corpus and the query have converted to lower cases. Unnecessary special characters like punctuation marks have been removed.
2. Tokenization has been applied to both the corpus and user query.

Building Keywords

Once the corpus has cleaned, list of keywords that our model will look for is built.

Building Intent Keywords

After the keywords list is completed, a dictionary that matches the user's query keywords has constructed to the required intents. We reformat the keywords in a special syntax that helps to get the required answer. We tried to design a common intent for a conversation that use common approach.

Defining Response Dictionary

We defined responses for each intent type as a key.

```
QUERY={'nagaa':[Akkam? Nagaa ooltee? Fayyaadhaa? Akkam bulte?'],
'yeroo':[yeroon meeqa?, sahatiin meeqa? sa'atii meeqatti dhufta?'],
'Walii gala':[Maqaan kee eenyu?, Eenyutu si uume?, Na jaallattaa?'],
}
```

Match Input to intent based on keyword and Response Selector

An input from the user is matched with the keywords dictionary using RegEx Search function.
The current intent gets selected if the match is found and the key to the responses has selected.

```
const reply = [["nagaa!", "fayyaa!", "galata waaqayyoo!", "Galata rabbii!","jirra","nagumaa"], [ "sa'a 6", "sa'a booda", "boru" ]]
```

Figure 2. A Sample response

```
function Chatbot(user input) returns response
Find the word i in input that has the intent
    if i exists
        Choose the response that matches input
        response Apply the transform in r to input
    else (no keyword applies)
        response Apply the transform for the NONE keyword to input
    return(response)
```

Figure 3. Rule based Algorithm

B. The Model Results

Chatbot system has significant applications in transforming the human being's lives. For the evaluation purpose, we collected corpus during the experiment phase, the Afaan Oromo language conversations of users of the total of 114 conversations and 30 users, 20 conversations were selected for the evaluation purpose. The model performs a good result by imitating a character user and system from the sample conversations.

Right responses are given to user to to make a small conversation. The work is analysed by human evaluators, ten (10) evaluators were selected to

evaluate the performance of the model and the average score being evaluated was 88%.

Figure 4. Afaan Oromo Rule Based Chatbot system User Interface

The above figure denotes the physical interface where a user asks a system to get his /her request.

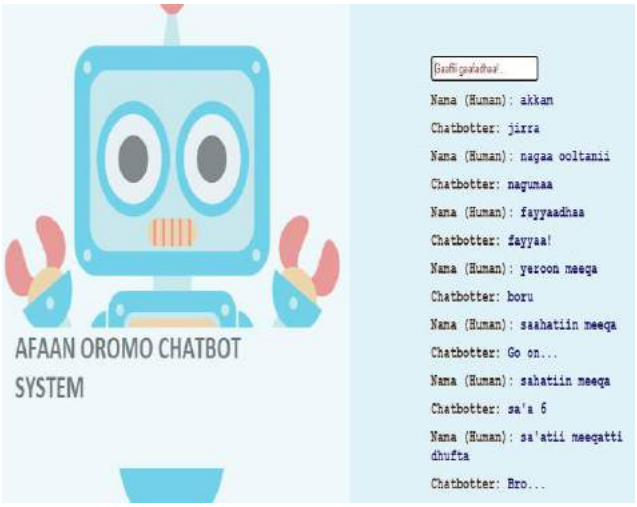


Figure 5 A Sample Conversations

A user is denoted by Nama (Human) and Here the user is denoted by “Human” and the chatbot is represented by “Chatbotter”.

IV. CONCLUSION AND FUTURE WORK

In any language, chatbots are user-friendly for any user who has an awareness of typing in the language they can understand on the desktop application and in the mobile. In this work, we designed a simple rule based Chatbot for Afaan Oromo Language. We only tested our model with 20 intents in this work for the evaluation purpose. From our result we saw designing and implementing a rule-based chatbot is a very challenging task since a rule based chatbot requires more intents.

In addition, the chatbot would severely be limited in terms of its conversational capabilities as it is near impossible to describe exactly how a user will interact with the bot. since Afaan Oromo has very complex structure, defining all rule was very challenging one. Another challenge was grammar error cannot be recognized.

We can make improvements to the developed chatbot to further enhance its affectivity since rule based has many limitations. First, AI-based Chatbots are a more practical solution for solving the limitations of rule based scenarios. So, we’ll build an AI-based Chatbot for this language.

V. REFERENCES

[1]. D’ÁVILA, THIAGO CARVALHO. KINO: AN APPROACH FOR RULE-BASED CHATBOT DEVELOPMENT, MONITORING ANDEVALUATION. 2018.

[2]. Jagdish Singh, Minnu Helen Joesph and Khurshid Begum Abdul Jabbar. "Rule-based chabot for student enquiries." International conference on computer vision and machine learning, 2019.

[3]. Martin, Daniel Jurafsky & James H. Speech and Language Processing. 2018.

[4]. O’Reilly, Amir Shevat. "Designing Bots." 2017.

[5]. Soufyane Ayanouz, Soufyane Ayanouz, Mohammed Benhmed. "A Smart Chatbot Architecture based NLP and Machine learning for health care assistance." Research Gate, April 2020.

[6]. Weizenbaum, J. "ELIZA – A computer program forthe study of natural language communication between manand machine." Communications of the ACM, 1996: 36–45.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

A New Class of Skew Normal Distribution : Tanh- Skew Normal Distribution and its Properties

Mahmoud R. Mahmoud[1], Hanan H. El-Damrawy[2], Hassan S. Abdalla*[3]
[1]Institute of Statistical Studies & Research, Cairo University, Egypt
[2]Mathematics Department, Faculty of Science, Tanta University, Tanta, Egypt.
[3]Mathematics Department, Faculty of Science, Tanta University, Tanta, Egypt.

ABSTRACT

In this article we suggest a new class of skew normal distribution. It will be referred to as Tanh skew-normal distribution, where (Tanh) is a hyperbolic tangent function; a class of skew-normal distribution is proposed by considering a new skew function, It is not a probability distribution function, some properties of this new class distribution have been investigated. Several properties of this distribution have been discussed; parameters estimation using moments, moment generating function, maximum likelihood method, and Fisher information matrix are obtained. A numerical experiment was performed to see the behavior of MLEs. Finally, we apply this model to a real data-set to show that the new class distribution can produce a better fit than other classical Skew normal.

Keywords : Skew-Normal Distribution, Tanh Skew-Normal Distribution, Moments, Maximum Likelihood Estimators, Fisher Information Matrix.

I. INTRODUCTION

Azzalini, (1985) has presented the skew-normal (SN) distribution, it consists of modifying the normal probability density function by multiplication with a skewing function. Azzalini stated that the formal, $H(x) = 2\,f(x)\,\mathrm{G}(x)$, is the density of the skew distribution (pdf), where f is the density of a variable symmetric around (0), and (G),is the (cdf) of another independent random variable. By combining different symmetric distributions (normal, t, logistic, uniform, double exponential, etc.), numerous families of skewed distributions may be generated. This form of distribution is called a skew-symmetric, and it has been used in many applications to analyzes the asymmetric behavior of empirical data sets from various research fields, for more details about it, see (Ma and Genton, (2004)). In recent years, (Arellano-Valle et al. (2004) and G´omez et al. (2006)) introduced a class of skew-normal distributions, which related to the model (SN) introduced by Azzalini, most of those classes include the normal distribution as a particular case and satisfy similar properties as the normal family. (Hutson and Mudholkar (2000)), presented a normal a symmetric family of distributions with a different structure of the class (SN) considered by Azzalini, (1985), which is called epsilon skew-normal and it is denoted by (ESN(ε)), ε > 0 represents the asymmetry parameter, so that (ESN (0)) corresponds to the normal distribution. (Huang and Chen (2007), Chakraborty, Hazarika, and Ali (2012)) investigated

the generalized skew symmetric distributions by introducing a skew function in place of cumulative distribution function (cdf), F(•), where a skew function G(•) is satisfying properties, $0 \le G(x) \le 1$, $G(x)+G(-x)=1$, on the other hand, several of authors have provided alternative methods of Azallin, such as, (Elal-Olivero (2010), El-Damrawy et al, (2013b), Mahmoud, et al. (2015), Hazarika, Shah, and Chakraborty, (2019), Kundu and Gupta (2010).

This paper aims to introduce a new class of skew-normal distribution. It will be referred to as Tanh skew-normal distribution, where (Tanh) is a hyperbolic tangent function; a new class of skew-normal distribution is proposed by considering a new skew function where the skew function is not a cumulative distribution function (CDF). We expect that the proposed this model may be better (at least in terms of model fitting) than another classical skew-normal in certain practical situations.

The remainder of this paper is organized as follows: We discuss some of its basic properties are investigated, in Section 2. In Section 3, we provide expansions for a new skew-normal distribution, cumulative and density functions. In Section 4, we present various properties of the new model such as, moment generating function, and moments. In Section 5, the maximum likelihood estimator of the parameters of our model and an application to a real data set are obtained. Concluding remarks are presented in Section 6.

II. BASIC PROPERTIES

In this section we discuss the basic properties of the new distribution

2.1 . Define of The skew function:

The skew function $G(\lambda x)$ is using the hyperbolic tangent function (Tanh) and it's express by,

$$G(\lambda x) = 0.5\left(1 + Tanh\left[\frac{\lambda x}{2}\right]\right) \tag{1}$$

Where λ , is skew parameter, and the hyperbolic tangent function (Tanh) is the expression by

$$\text{Tanh}\left[\frac{\lambda x}{2}\right] = \frac{e^{\lambda x/2} - e^{-\lambda x/2}}{e^{\lambda x/2} + e^{-\lambda x/2}} \tag{2}$$

Substituting from (2) into (1) we get the skew function $G(\lambda x)$

$$G(\lambda x) = 0.5\left\{\frac{2}{1+e^{-\lambda x}}\right\} = \frac{1}{1+e^{-\lambda x}} = \left(1+e^{-\lambda x}\right)^{-1} \tag{3}$$

By using the Taylor series expansion for $(1+x)^{-1} = \sum_{k=0}^{\infty}(-1)^k x^k$,We get the skew function, $G(\lambda x)$ is,

$$G(\lambda x) = \begin{cases} \sum_{k=0}^{\infty}(-1)^k e^{-\lambda x}, & x \ge 0 \\ \sum_{k=0}^{\infty}(-1)^k e^{(k+1)\lambda x}, & x < 0 \end{cases} \tag{4}$$

Where , $\lambda > 0$

Similarly if $\lambda < 0$, the skew function is,

$$G(-\lambda x) = \begin{cases} \sum_{k=0}^{\infty}(-1)^k e^{(k+1)\lambda x}, & x > 0 \\ \sum_{k=0}^{\infty}(-1)^k e^{-\lambda x}, & x \le 0 \end{cases} \tag{5}$$

© 2020 IJSRST | Print ISSN: 2395-6011 | Online ISSN: 2395-602X | Published : 18/12/2020-31-44

2.2 . Properties of the skew function

a) The properties of the skew function are giving by,

$$G(\lambda x) + G(-\lambda x) = 1, 0 \leq G(\lambda x) \leq 1 \quad (6)$$

b) Plots of the skew function

The skew function of $G(\lambda x)$ for different choices of the parameter λ is plotted in Figure 1.

$$G(\lambda x) = 0.5(1 + Tanh[\lambda x/2])$$

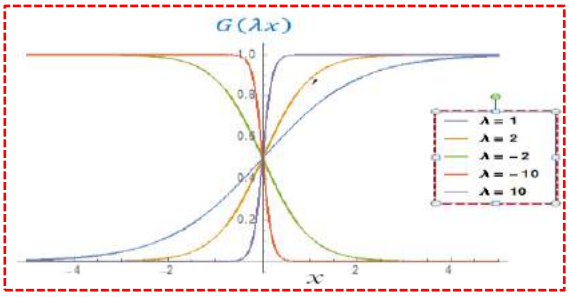


Figure 1.illustrates the shape of theskew function ($G(\lambda x)$for $\lambda = 1,\ 2, -2, 10, -10$

3. Expansions of The density and the cumulative distribution function

In this section, we introduce the probability density function and the cumulative distribution function of thestandardtanh skew- normal distribution.

3-1 Expansions of The Probability density function

The probability density function of thestandardtanh skew- normal distribution($TSN(\lambda)$) is constructed using the formula ,

$$f(x, \lambda) = 2\phi(x)\, G(\lambda x), \quad -\infty < x < \infty \quad (7)$$

Where λ, a real number, is the skewness parameter,$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, is the probability density function of standard normal distribution and $G(\lambda x)(not\ cdf)$ is the skew function of definition its above equation (1,4). Substituting, $\phi(x), and\ G(\lambda x)$into formula (7), we get the probability density function,

$$f(x, \lambda) = \begin{cases} \frac{2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left(\sum_{k=0}^{\infty} (-1)^k e^{-\lambda x} \right), & x > 0 \\ \frac{2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left(\sum_{k=0}^{\infty} (-1)^k e^{(k+1)\lambda x} \right), & x < 0 \end{cases} \quad (8)$$

Where,$\lambda > 0$

Similarly $\boldsymbol{\lambda} < 0$, the probability density function is

$$f(x, \lambda) = \begin{cases} \frac{2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left(\sum_{k=0}^{\infty} (-1)^k e^{(k+1)\lambda x} \right), & x > 0 \\ \frac{2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left(\sum_{k=0}^{\infty} (-1)^k e^{-\lambda x} \right), & x < 0 \end{cases} \quad (9)$$

3.1.1 Properties of the probability density function

The properties of the density function of the standard tanh skew- normaldistribution($TSN(\lambda)$)are giving by,

a) $f(x,\lambda) \geq 0$

b) $\int_{-\infty}^{\infty} f(x,\lambda)\, dx = 1$

c) if $\lambda = 0$, then we get the standard normal distribution and is given, $f(x,\lambda) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

d) If $X \tilde{} TSN(X,\lambda)$ then $-X \tilde{} TSN(X,-\lambda)$

e) Plots of the probability density function : the plots of the probability density function of $TSN(\lambda)$ distribution for different choices of the parameter λ, is plotted in Figure 2.

$$f(x,\lambda) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left(1 + \text{Tanh}\left[\frac{\lambda x}{2}\right]\right), -\infty < x < \infty, \quad \lambda \in R$$

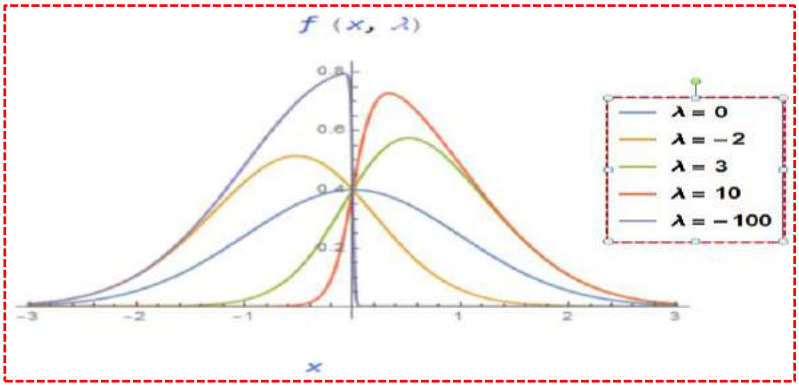


Figure2. illustrates the shape of the probability density function for λ= 0, -2, 3, 10, -100.

Remark 1.

from the Figure2 , when ,$\lambda \to \pm\infty$ the probability density function of the standard Tanh skew- normal distribution(*TSN(λ)*) converges to the half normal density function.

Remark 2.

Throughout the rest of this paper (unless otherwise stated), we shall assume that $\lambda > 0$, since the corresponding results for $\lambda < 0$, can be obtained using the fact that *(–X)* has the density function

$$f(x) = 2\phi(x)\, G(-\lambda x).$$

3.2 Expansions of the cumulative distribution function:

In this subsection, we introduce the expansion forms for thecumulative distribution function(cdf) for the standardTanhskew normal distribution , we write it's at following ,

$$F(x) = \int_{-\infty}^{x} f(x)\, dx = \int_{0}^{x} f(x)\, dx + \int_{-\infty}^{0} f(x)\, dx$$

where $F(x)$ is definite the cumulative function.

1) If$x \geq 0$, $\lambda > 0$, and writing ,$F_1(x) = A + B$

$$A = \int_{0}^{x} f(x)\, dx, \qquad B = \int_{-\infty}^{0} f(x)\, dx$$

Substituting about $f(x)$ in equation (8), we getting,

$$A = \int_{0}^{x} \frac{2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left(\sum_{k=0}^{\infty} (-1)^k e^{-k\lambda x}\right) dx = \sqrt{\frac{2}{\pi}} \sum_{k=0}^{\infty} (-1)^k \int_{0}^{x} e^{-\frac{x^2}{2} - k\lambda x}\, dx\,,$$

$$A = \sum_{k=0}^{\infty} (-1)^k \; e^{(\frac{k^2\lambda^2}{2})} \left\{ \; erf\left[\frac{1}{\sqrt{2}} \; (x + k\lambda)\right] - erf\left[\frac{1}{\sqrt{2}} \; (k\lambda)\right]\right\}$$

Similarly, $B = \int_{-\infty}^{0} f(x)\, dxB = \int_{-\infty}^{0} f(x)\, dx = \frac{2}{\sqrt{2\pi}} \sum_{k=0}^{\infty} (-1)^k \int_{-\infty}^{0} e^{-\frac{1}{2}x^2 + (k+1)\lambda x} dx$

$$B = \sum_{k=0}^{\infty} (-1)^k \; e^{\frac{(k+1)^2\lambda^2}{2}} \left[erfc\left[\frac{(k+1)\lambda}{\sqrt{2}}\right]\right]$$

From above equations (A,B), we get the cumulative function, $\lambda > 0$

$$F_1(x) = \begin{cases} \sum_{k=0}^{\infty} (-1)^k \; e^{\left(\frac{k^2\lambda^2}{2}\right)} \left\{erf\left[\frac{x + k\lambda}{\sqrt{2}}\right] - erf\left[\frac{k\lambda}{\sqrt{2}}\right]\right\} \\ + \; \sum_{k=0}^{\infty} (-1)^k \; e^{\frac{(k+1)^2\lambda^2}{2}} \; erfc\left[\frac{(k+1)\lambda}{\sqrt{2}}\right], \qquad x \geq 0, \end{cases} \tag{10}$$

2) If $\boldsymbol{x} < 0, \;\; \lambda > 0,$ the cumulative distribution function is,

$$F_2(x) \;\; = \int_{-\infty}^{x} f(x)\, dx == \frac{2}{\sqrt{2\pi}} \sum_{k=0}^{\infty} (-1)^k \int_{-\infty}^{x} e^{\frac{-x^2}{2} + (k+1)\lambda x} \; dx$$

$$F_2(x) = \sum_{k=0}^{\infty} (-1)^k \; e^{\frac{(k+1)^2\lambda^2}{2}} \left\{erfc\left[\frac{\lambda(k+1) \; - \; x}{\sqrt{2}}\right]\right\}, x < 0 \tag{11}$$

Since, from equations (10),(11), the cumulative distribution function is,$F(x) =$

$$F_1(x) + F_2(x) = \begin{cases} \sum_{k=0}^{\infty} (-1)^k \; e^{\left(\frac{k^2\lambda^2}{2}\right)} \left\{erf\left[\frac{(x+k\lambda)}{\sqrt{2}}\right] - erf\left[\frac{(k\lambda)}{\sqrt{2}}\right]\right\} \\ + \sum_{k=0}^{\infty} (-1)^k \; e^{\frac{(k+1)^2\lambda^2}{2}} \; erfc\left[\frac{(k+1)\lambda}{\sqrt{2}}\right], \qquad x \geq 0 \\ \sum_{k=0}^{\infty} (-1)^k \; e^{\frac{(k+1)^2\lambda^2}{2}} \; erfc\left[\frac{(k+1)\lambda - \; x}{\sqrt{2}}\right], \; x < 0 \end{cases} \tag{12}$$

Similarly, if $\lambda < 0,$ By using the dinsty function in equation (9) we getting,

The Cumulative distribution function is,

$$F(x) = \left\{ \begin{array}{c} \sum_{k=0}^{\infty} (-1)^k \; e^{\frac{(k+1)^2\lambda^2}{2}} \left\{erf\left[\frac{1}{\sqrt{2}} \; x - \frac{(k+1)\lambda}{\sqrt{2}}\right] + erf\left[\frac{(k+1)\lambda}{\sqrt{2}}\right]\right\} \\ + \; \sum_{k=0}^{\infty} (-1)^k \, e^{\left(\frac{k^2\lambda^2}{2}\right)} \left\{1 + erf\left[\frac{k\lambda}{\sqrt{2}}\right]\right\}, \qquad x \geq 0 \\ \sum_{k=0}^{\infty} (-1)^k \; e^{(\frac{k^2\lambda^2}{2})} \left\{1 + erf\left[\frac{1}{\sqrt{2}} \; x + \frac{k\lambda}{\sqrt{2}}\right]\right\}, \qquad x < 0 \end{array} \right\} \tag{13}$$

where (*erf*) is denoted error function, and its donation, (1) $erf\,(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-x^2}$, (erfc) is denoted complement error function, (2) $erfc(x) = 1 - erf\,(x)$)

3..2.1 plots of the cumulative distribution function

the plots of The cumulative distribution function with different choices parameter $\lambda,$ is shown in figure 4.

© 2020 IJSRST | Print ISSN: 2395-6011 | Online ISSN: 2395-602X | Published : 18/12/2020-31-44

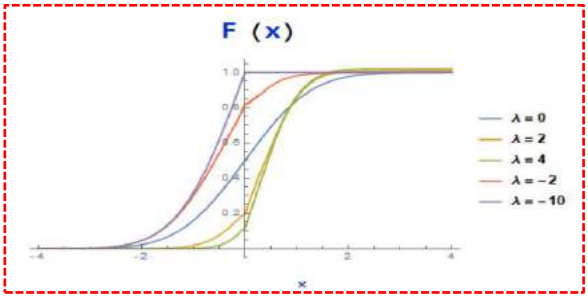


Figure 3. Plots of the cumulative distribution function (cdf) of *TSN(λ))* ,λ= 0, 2, 4, -2, -10.

3.2.2. Properties of the Cumulative distribution function

a) if $\lambda = 0$, The Cumulative distribution function is stander normal distribution
b) The skewness of the distribution($TSN(\lambda)$) increases as the value of λ increases in absolute.
c) When $\lambda \to \pm\infty$ The Cumulative distribution function of the standard tanh skew- normal distribution (*TSN (λ))* converges to half normal density function.
d) $0 \leq F(x) \leq 1$

4. Statistical Properties

In this section, Let us view different moments of $TSN(0,1,\lambda)$distribution. By using the moment , we can study some of the most important characteristics and features of a distribution, such as moment generating function, characteristic function and moments.

4.1. Moment Generating Function

If X has the $TSN(\lambda)$distribution, *and* $\lambda > 0$ then the mgf is:

$$M_x(t) = \sum_{k=0}^{\infty} (-1)^k \left(e^{\frac{1}{2}(k\lambda - t)^2} \left[erfc\left(\frac{1}{\sqrt{2}} (k\lambda - t) \right) \right] + e^{\frac{1}{2}((k+1)\lambda + t)^2} \left[erfc\left(\frac{1}{\sqrt{2}} ((k+1)\lambda + t) \right) \right] \right) \quad (14)$$

4.2. The characteristic function

The characteristic function of $X$is:

$$E[\exp(itX)] = \sum_{k=0}^{\infty} (-1)^k \left\{ e^{\frac{[(k+1)\lambda + it]^2}{2}} \operatorname{erfc}\left[\frac{(k+1)\lambda + it}{\sqrt{2}} \right] - e^{\frac{1}{2}(k\lambda - it)^2} \left(\operatorname{erfc}\left[\frac{k\lambda - it}{\sqrt{2}} \right] \right) \right\} \quad (15)$$

Where , $\lambda > 0, i = \sqrt{-1}$

4.3 General Moments

If $X \tilde{} TSN(0,\lambda)$ is a random variable, the r-th moments of *X, is defined as*

$$\acute{\mu}_r = E(X^r) = \int_{-\infty}^{\infty} x^r f(x)\, dx,$$

Using these representations and properties in Sections 2.3 of (Prudnikov **et** *al.*[**13,** volume 1]), one can obtain,

$$E(X^r) = \lambda \sum_{k=0}^{\infty} (-1)^k \left\{ \frac{\partial^r}{\partial q^r} e^{\frac{(k+1)^2\lambda^2}{2}} erfc\left[\frac{(k+1)\lambda}{\sqrt{2}} \right] + (-1)^r \frac{\partial^r}{\partial s^r} e^{\frac{k^2\lambda^2}{2}} erfc\left[\frac{k\lambda}{\sqrt{2}} \right] \right\} \quad (16)$$

Where $q = \lambda(k+1),\ s = \lambda k$

4.2.1 Mean

For r= 1 Equation (16), yields the mean $(E(X) = \acute{\mu}_1)$ of $TSN(\lambda)$that is given by:

$$\acute{\mu}_1 = \mathrm{E(X)} = \sum_{k=0}^{\infty} (-1)^k \left\{ \lambda(k+1)\, e^{\frac{(k+1)^2\lambda^2}{2}} \mathrm{erfc}\left[\frac{(k+1)\lambda}{\sqrt{2}}\right] - k\lambda e^{\frac{k^2\lambda^2}{2}} \left(\mathrm{erfc}\left[\frac{k\lambda}{\sqrt{2}}\right]\right)\right\} \qquad (17)$$

We derive the moment generating function at $t = 0$ we can getting the four first moments by using mgf$\left(M_x(t)\right)$, and formula $\acute{\mu}_r$ is

$$\acute{\mu}_r = \frac{d^r}{dt^r}\left(M_x(t)\right), \qquad t = 0$$

$$2)\ E(X^2) = \sum_{k=0}^{\infty} (-1)^k \left\{ \begin{array}{c} -\sqrt{\frac{2}{\pi}} k\lambda + (k^2\lambda^2+1) e^{\frac{k^2\lambda^2}{2}} erfc\left[\frac{k\lambda}{\sqrt{2}}\right] - \sqrt{\frac{2}{\pi}}\lambda(k+1) \\ +((k+1)^2\lambda^2+1) e^{\frac{(k+1)^2\lambda^2}{2}} \left(erfc\left[\frac{(k+1)\lambda}{\sqrt{2}}\right]\right) \end{array} \right\} \qquad (18)$$

$$3)\ E(X^3) = \sum_{k=0}^{\infty} (-1)^k \left\{ \begin{array}{c} -\sqrt{\frac{2}{\pi}}(2k+1)\lambda^2 - k\lambda(3+k^2\lambda^2) e^{0.5k^2\lambda^2} erfc\left[\frac{1}{\sqrt{2}}\, k\lambda\right] \\ +\lambda(k+1)((k+1)^2\lambda^2+3)\, e^{\frac{1}{2}(k+1)^2\lambda^2} erfc\left[\frac{1}{\sqrt{2}}\left((k+1)\lambda\right)\right] \end{array} \right\} \qquad (19)$$

$$4)\ E(X^4) = \sum_{k=0}^{\infty} (-1)^k \left\{ \begin{array}{c} -\sqrt{\frac{2}{\pi}}\lambda(k+1)(5+(k+1)^2\lambda^2) - \sqrt{\frac{2}{\pi}} k\lambda(5+k^2\lambda^2) \\ +(k^4\lambda^4+6k^2\lambda^2+3) e^{\frac{k^2\lambda^2}{2}} erfc\left[\frac{k\lambda}{\sqrt{2}}\right] \\ +((k+1)^4\lambda^4+6(k+1)^2\lambda^2+3) e^{\frac{(k+1)^2\lambda^2}{2}} erfc\left[\frac{(k+1)\lambda}{\sqrt{2}}\right] \end{array} \right\} \qquad (20)$$

4.2.2 Central Moments

The central moments (μ_r) of Xcan be calculated as,

$$\mu_r = E[(X - \acute{\mu}_1)^r] \quad , \text{ where } \acute{\mu}_1 \text{ is the mean of } X,$$

$$\mu_r = \int_{-\infty}^{\infty} (x - \acute{\mu}_1)^r f(x)\, dx, \qquad \lambda > 0$$

By using binomial expansion

$$(a-b)^m = \sum_{j=0}^{m} (-1)^j \binom{m}{j} b^j a^{m-j}$$

$$\mu_r = \sum_{j=0}^{r} (-1)^j \binom{r}{j} (\acute{\mu}_1)^j \int_{-\infty}^{\infty} x^{r-j} f(x)\, dx, \qquad \lambda > 0$$

$$\mu_r = \textstyle\sum_{j=0}^{r} (-1)^j \binom{r}{j} (\acute{\mu}_1)^j \acute{\mu}_{r-j} \qquad (21)$$

Where $\mu_1 = \acute{\mu}_1$.Then, the variance of $TSN(\lambda)$ distribution is given by:$\mu_2 = \acute{\mu}_2 - \acute{\mu}_1^2$, $\mu_3 = \acute{\mu}_3 - 3\,\acute{\mu}_2\,\acute{\mu}_1 + 2\,\acute{\mu}_{1,}^2$, $\mu_4 = \acute{\mu}_4 - 4\,\acute{\mu}_3 + 6\,\acute{\mu}_2\,\acute{\mu}_1^2 - 3\,\acute{\mu}_{1,}^4$, etc. Also, the skewness$\gamma_1 = \mu_3 / \mu_2^{\frac{3}{2}}$, and kurtosis $\gamma_2 = \mu_4 / \mu_2^2$ follow from the second, third and fourth moments.

4.2.3 Mean Deviation

Let Xbe a random variable that follows $TSN(\lambda)$ distribution with median m and

mean μ. In this subsection, we inferred the mean deviation from the mean and the median.

1) The form of the mean deviation from the mean of the $TSN(\lambda)$ distribution is,

$$E(|X-\mu|) = \int_{-\infty}^{\infty} |X-\mu|\, f(x)\, dx,$$

$$= \int_{-\infty}^{\mu} (\mu - x)\, f(x)\, dx \;\; + \;\; \int_{\mu}^{\infty} (x-\mu)\, f(x)\, dx$$

$$E(|X-\mu|) = \; 2\,\mu\, F(\mu) \;\; -2\int_{-\infty}^{\mu} x f(x)\, dx\,,$$

- If $\mu \geq 0$

$$E(|X-\mu|) = \; 2\,\mu \left\{ \begin{array}{l} \sum_{k=0}^{\infty} (-1)^k \; e^{\left(\frac{k^2\lambda^2}{2}\right)} \left\{ erf\left[\frac{\mu + k\lambda}{\sqrt{2}}\right] - erf\left[\frac{k\lambda}{\sqrt{2}}\right]\right\} \\ + \sum_{k=0}^{\infty} (-1)^k \; e^{\frac{(k+1)^2\lambda^2}{2}} erfc\left[\frac{(k+1)\lambda}{\sqrt{2}}\right], \end{array} \right\}$$

$$- 2\sum_{k=0}^{\infty} (-1)^k \left\{ (k+1)\lambda\, e^{\frac{(k+1)^2\lambda^2}{2}} erfc\left[\frac{(k+1)\lambda}{\sqrt{2}}\right] - 1 \right\} - 2\int_{0}^{\mu} x f(x)\, dx\,,$$

- If $\mu < 0$

$$E(|X-\mu|) \;\; = \;\; 2\,\mu \sum_{k=0}^{\infty} (-1)^k \; e^{\frac{(k+1)^2\lambda^2}{2}} erfc\left[\frac{(k+1)\lambda \; - \; \mu}{\sqrt{2}}\right] - 2\int_{-\infty}^{\mu} x f(x)\, dx\,,$$

3) The form of the mean deviation from the median of the $TSN(\lambda)$ distribution

$$E(|X-m|) = \int_{-\infty}^{\infty} |X-m|\, f(x)\, dx,$$

The expression for $E(|X-m|)$is the same with μ replaced by m.

4.4 Location Scale Extension:

The location and scale extension of $TSN(\mu,\sigma,\lambda)$distribution is as follows. If Z ~ $TSN(\mu,\sigma,\lambda)$ then $X= Z\,\sigma + \mu$, is said to be the location (μ) and scale (σ) extension of Z and has the density function is given by

$$f(z,\lambda) = 2\phi\left(\tfrac{x-\mu}{\sigma}\right) G\left(\tfrac{\lambda(x-\mu)}{\sigma}\right),$$

where, (z, μ,σ,λ) εR, and $\sigma > 0$ We denote it by $X\sim\ TSN(\mu,\sigma,\lambda)$

5. Parameter Estimation and Applications

5.1 Likelihood function and maximum likelihood estimates.

In this section, the ML method is considered to estimate the parameters of $TSN(\mu,\sigma,\lambda)$ distribution. Let $(x_1, x_2, \ldots\ldots x_n)$ be a random sample with size n from the $TSN(\mu,\sigma,\lambda)$, with *pdf*by $f(x_i,\mu,\sigma,\lambda) = 2\phi\left(\frac{x_i-\mu}{\sigma}\right) G\left(\frac{\lambda(x_i-\mu)}{\sigma}\right)$, Also, we assume that$\Theta = (\mu,\sigma,\lambda)^{\tau}$is the $(r*1)$unknown parameter vectors, the log-likelihood function is defined by:

$$l(\Theta) = \text{Log}\, \mathrm{L}_n(\Theta/x_1, x_2, \ldots\ldots, x_n) = \text{Log}\, \mathrm{L}_n(\Theta/x_i))$$

Where, $\mathrm{L}_n(\Theta/x_i) = \prod_{i=1}^{n} f(x_i,\Theta)$ is the likelihood function.We can derive the likelihood function of $TSN(\mu,\sigma,\lambda)$ distribution as,

$$\mathrm{L}_n(\Theta/x_i) = \prod_{i=1}^{n} 2\phi\left(\frac{x_i-\mu}{\sigma}\right) G\left(\frac{\lambda(x_i-\mu)}{\sigma}\right)$$

Where,$\phi\left(\frac{x_i-\mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}$,,$-\infty < \; x < \infty, \;\; \mu \in R\,, \;\; \sigma > 0$

is the probability density function of normal distribution and $G(\lambda x)$ is the skew function

$$G(\lambda x) = 0.5\left(1 + \text{Tanh}\left[\frac{\lambda x}{2}\right]\right) = \frac{1}{1 + e^{-\lambda\left(\frac{x_i-\mu}{\sigma}\right)}} \mu \epsilon R\,, \quad \sigma > 0, \;\; \lambda \epsilon R$$

Where, λ, is skew parameter

$$\text{L}_n(\mu, \sigma, \lambda / x_i) = \prod_{i=1}^{n} \frac{2}{\sigma\sqrt{2\pi}} \frac{\text{e}^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}}{\left[1 + e^{-\frac{\lambda(x_i-\mu)}{\sigma}}\right]}$$

Thus, the log-likelihood function of $TSN(\mu, \sigma, \lambda)$ distribution is obtained as:

$$l(\Theta) = \text{Log L}_n(\mu, \sigma, \lambda | x_i) = \text{Log}\left[\frac{2}{\sigma\sqrt{2\pi}}\right]^n - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2 - \sum_{i=1}^{n}\text{Log}\left[1 + e^{-\frac{\lambda(x_i-\mu)}{\sigma}}\right] \quad (22)$$

$$l(\Theta) = \text{n Log (2)} - \text{n Log}\,\sigma - \frac{n}{2}Log\,(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2 - \sum_{i=1}^{n}\text{Log}\left[1 + e^{-\frac{\lambda(x_i-\mu)}{\sigma}}\right] \quad (23)$$

The first derivatives of Equation (23) with respect to μ, σ,and λ respectively is given by:

$$\frac{\partial l(\Theta)}{\partial \mu} = \frac{\partial l(\mu, \sigma, \lambda)}{\partial \mu} = \frac{1}{\sigma}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right) - \frac{\lambda}{\sigma} \quad (24)$$

$$\frac{\partial l(\Theta)}{\partial \sigma} = \frac{\partial l(\mu, \sigma, \lambda)}{\partial \sigma} = \frac{-n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2 - \frac{\lambda}{\sigma}\sum_{i=1}^{n}\frac{\frac{(x_i-\mu)}{\sigma}e^{-\frac{\lambda(x_i-\mu)}{\sigma}}}{\left[1 + e^{-\frac{\lambda(x_i-\mu)}{\sigma}}\right]} \quad (25)$$

$$\frac{\partial l(\Theta)}{\partial \lambda} = \frac{\partial l(\mu, \sigma, \lambda)}{\partial \lambda} = \sum_{i=1}^{n}\frac{\frac{(x_i-\mu)}{\sigma}e^{-\frac{\lambda(x_i-\mu)}{\sigma}}}{\left[1 + e^{-\frac{\lambda(x_i-\mu)}{\sigma}}\right]} \quad (26)$$

Equate the Equations (24)-(26) to zero and solving them simultaneously yield the maximum likelihood estimators (MLEs) of $TSN(\mu, \sigma, \lambda)$ distribution parameters, then, from (24, 25, and 26), we get

$$\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right) = \lambda\sum_{i=1}^{n}\frac{e^{-\frac{\lambda(x_i-\mu)}{\sigma}}}{\left[1 + e^{-\frac{\lambda(x_i-\mu)}{\sigma}}\right]} \quad (27)$$

$$\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2 - \lambda\sum_{i=1}^{n}\frac{\frac{(x_i-\mu)}{\sigma}e^{-\frac{\lambda(x_i-\mu)}{\sigma}}}{\left[1 + e^{-\frac{\lambda(x_i-\mu)}{\sigma}}\right]} = n \quad (28)$$

$$\sum_{i=1}^{n}\frac{\frac{(x_i-\mu)}{\sigma}e^{-\frac{\lambda(x_i-\mu)}{\sigma}}}{\left[1 + e^{-\frac{\lambda(x_i-\mu)}{\sigma}}\right]} = 0 \quad (29)$$

Clearly, these equations are not in explicit form, the solutions can be found by using a numerical method such as the Newton-Raphson procedure to obtain the MLEs of the parameters μ, σ,and λ To obtain the asymptotic confidence intervals (CIs) for the parameters of the $TSN(\mu, \sigma, \lambda)$ distribution, the 3×3 $I_n(\Theta) = I(\mu, \sigma, \lambda)$is required. Under certain regularity conditions, the MLEs asymptotically have multivariate normal distribution with mean vector $(\underset{\sim}{\Theta} = 0,0,0)$ and variance- covariance matrix, which is given by the inverse of Fisher information:

$$I_n{}^{-1}\left(\widehat{\Theta}\right) = I_n{}^{-1}\left(\hat{\mu}, \hat{\sigma}, \hat{\lambda}\right) (30)$$

for more details about asymptotic confidence intervals. The $I_n(\Theta)$depends on Θ, the observed Fisher information matrix $I_n\left(\widehat{\Theta}\right)$May be used instead of the $I_n(\Theta)$ the estimation of the variance of

$$I_n\left(\widehat{\Theta}\right) = I_n(\,\hat{\mu}, \hat{\sigma}, \hat{\lambda})$$

$$I_n(\widehat{\Theta}) = -\begin{bmatrix} I_{\mu\mu} & I_{\mu\sigma} & I_{\mu\lambda} \\ I_{\sigma\mu} & I_{\sigma\sigma} & I_{\sigma\lambda} \\ I_{\lambda\mu} & I_{\lambda\sigma} & I_{\lambda\lambda} \end{bmatrix}_{(\widehat{\mu}=\mu,\widehat{\sigma}=\sigma,\widehat{\lambda}=\lambda)} \quad (31)$$

Where, $I_{ij} = E\left[\frac{\partial^2 Log\,\ell}{\partial\theta_i\partial\theta_j}\right]$regrettably, the accurate mathematical expressionsfor the above expectation are very hard to obtain. Therefore, the observed Fisher information matrix is given by which is obtained by$I_{ij} = \left[\frac{\partial^2 Log\,\ell}{\partial\theta_i\partial\theta_j}\right]$ dropping the expectation on operation The elements of the $I_n(\widehat{\Theta})$are given by the following equations,

$$I_{\mu\mu} = E\left[\frac{\partial^2}{\partial\mu^2}\text{Log}\,L_n(\mu,\sigma,\lambda)\right] = \frac{-n}{\sigma^2} - \frac{\lambda^2}{\sigma^2}\sum_{i=1}^{n} E\left(\frac{e^{-\frac{\lambda(x_i-\mu)}{\sigma}}}{\left[1+e^{-\frac{\lambda(x_i-\mu)}{\sigma}}\right]}\right) + \frac{\lambda^2}{\sigma^2}\sum_{i=1}^{n} E\left(\frac{e^{-\frac{\lambda(x_i-\mu)}{\sigma}}}{\left[1+e^{-\frac{\lambda(x_i-\mu)}{\sigma}}\right]}\right)^2 \quad (32)$$

$$I_{\sigma\sigma} = \left(\begin{array}{l} \frac{n}{\sigma^2} - \frac{2}{\sigma^2}\sum_{i=1}^{n} E\left(\frac{x_i-\mu}{\sigma}\right)^2 + \frac{2\lambda}{\sigma^2}\sum_{i=1}^{n} E\left(\frac{\frac{(x_i-\mu)}{\sigma}e^{-\frac{\lambda(x_i-\mu)}{\sigma}}}{\left[1+e^{-\frac{\lambda(x_i-\mu)}{\sigma}}\right]}\right) - \frac{\lambda^2}{\sigma^2}\sum_{i=1}^{n} E\left(\frac{\left(\frac{x_i-\mu}{\sigma}\right)^2 e^{-\frac{\lambda(x_i-\mu)}{\sigma}}}{\left[1+e^{-\frac{\lambda(x_i-\mu)}{\sigma}}\right]}\right) \\ \qquad + \frac{\lambda^2}{\sigma^2}\sum_{i=1}^{n} E\left(\frac{\left(\frac{x_i-\mu}{\sigma}\right)e^{-\frac{\lambda(x_i-\mu)}{\sigma}}}{\left[1+e^{-\frac{\lambda(x_i-\mu)}{\sigma}}\right]}\right)^2 \end{array}\right) \quad (33)$$

$$I_{\lambda\lambda} = -\sum_{i=1}^{n} E\left(\frac{\left(\frac{x_i-\mu}{\sigma}\right)^2 e^{-\frac{\lambda(x_i-\mu)}{\sigma}}}{\left[1+e^{-\frac{\lambda(x_i-\mu)}{\sigma}}\right]}\right) + \sum_{i=1}^{n} E\left(\left(\frac{x_i-\mu}{\sigma}\right)^2\left(\frac{e^{-\frac{\lambda(x_i-\mu)}{\sigma}}}{\left[1+e^{-\frac{\lambda(x_i-\mu)}{\sigma}}\right]}\right)^2\right) \quad (34)$$

$$\begin{aligned} I_{\mu\sigma} = I_{\sigma\mu} &= E\left[\frac{\partial}{\partial\mu}\left(\frac{\partial}{\partial\sigma}\text{Log}\,L_n(\mu,\sigma,\lambda)\right)\right] \\ &= -\frac{2}{\sigma^2}\sum_{i=1}^{n} E\left(\frac{x_i-\mu}{\sigma}\right) + \frac{\lambda}{\sigma^2}\sum_{i=1}^{n} E\left(\frac{e^{-\frac{\lambda(x_i-\mu)}{\sigma}}}{1+e^{-\frac{\lambda(x_i-\mu)}{\sigma}}}\right) - \frac{\lambda^2}{\sigma^2}\sum_{i=1}^{n} E\left(\frac{\frac{(x_i-\mu)}{\sigma}e^{-\frac{\lambda(x_i-\mu)}{\sigma}}}{1+e^{-\frac{\lambda(x_i-\mu)}{\sigma}}}\right) \\ &\quad + \frac{\lambda^2}{\sigma^2}\sum_{i=1}^{n} E\left(\left(\frac{x_i-\mu}{\sigma}\right)\left(\frac{e^{-\frac{\lambda(x_i-\mu)}{\sigma}}}{1+e^{-\frac{\lambda(x_i-\mu)}{\sigma}}}\right)^2\right) \end{aligned} \quad (35)$$

$$\begin{aligned} I_{\mu\lambda} = I_{\lambda\mu} &= E\left[\frac{\partial}{\partial\mu}\left(\frac{\partial}{\partial\lambda}\text{Log}\,L_n(\mu,\sigma,\lambda)\right)\right] \\ &= \frac{-1}{\sigma}\sum_{i=1}^{n} E\left(\frac{e^{-\frac{\lambda(x_i-\mu)}{\sigma}}}{1+e^{-\frac{\lambda(x_i-\mu)}{\sigma}}}\right) \\ &\quad + \frac{\lambda}{\sigma}\sum_{i=1}^{n} E\left(\frac{\frac{(x_i-\mu)}{\sigma}e^{-\frac{\lambda(x_i-\mu)}{\sigma}}}{1+e^{-\frac{\lambda(x_i-\mu)}{\sigma}}}\right) + \frac{\lambda}{\sigma}\sum_{i=1}^{n} E\left(\left(\frac{x_i-\mu}{\sigma}\right)\left(\frac{e^{-\frac{\lambda(x_i-\mu)}{\sigma}}}{1+e^{-\frac{\lambda(x_i-\mu)}{\sigma}}}\right)^2\right) \end{aligned} \quad (36)$$

$$\boldsymbol{I_{\sigma\lambda}} = \boldsymbol{I_{\lambda\sigma}} = E\left[\frac{\partial}{\partial\sigma}\left(\frac{\partial}{\partial\lambda}\text{Log}\, L_n(\mu,\sigma,\lambda)\right)\right]$$
$$= \frac{-1}{\sigma}\sum_{i=1}^{n} E\left\{\frac{\left(\frac{x_i-\mu}{\sigma}\right)e^{-\frac{\lambda(x_i-\mu)}{\sigma}}}{\left[1+e^{-\frac{\lambda(x_i-\mu)}{\sigma}}\right]}\right\} + \frac{\lambda}{\sigma}\sum_{i=1}^{n} E\left\{\frac{\left(\frac{x_i-\mu}{\sigma}\right)^2 e^{-\frac{\lambda(x_i-\mu)}{\sigma}}}{\left[1+e^{-\frac{\lambda(x_i-\mu)}{\sigma}}\right]}\right\} \quad (37)$$
$$+ \frac{\lambda}{\sigma}\sum_{i=1}^{n} E\left(\frac{\left(\frac{x_i-\mu}{\sigma}\right)e^{-\frac{\lambda(x_i-\mu)}{\sigma}}}{\left[1+e^{-\frac{\lambda(x_i-\mu)}{\sigma}}\right]}\right)^2$$

The approximate (1–δ)100% CIs of the parameters of $TSN(\mu,\sigma,\lambda)$ are respectively, given by: $\hat{\mu} \pm Z_{\frac{\delta}{2}}\sqrt{V(\hat{\mu})}$, $\hat{\sigma} \pm Z_{\frac{\delta}{2}}\sqrt{V(\hat{\sigma})}$, and $\hat{\lambda} \pm Z_{\frac{\delta}{2}}\sqrt{V(\hat{\lambda})}$ where, $V(\hat{\mu})$, $V(\hat{\sigma})$ and $V(\hat{\lambda})$ are the variances of $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\lambda}$ which are given by the diagonal elements of $I_n^{-1}(\hat{\Theta}) = I_n^{-1}(\hat{\mu},\hat{\sigma},\hat{\lambda})$ and $Z_{\frac{\delta}{2}}$ is the upper(δ / 2) percentile of the standard normal distribution.

5.2. Real Life Applications

In this section, we illustrate an application of the TSN(μ,σ,λ) distribution on the skew data set, for example, the data set is the white cells count (WCC) of 202 Australian athletes, given in (Cook and Weisberg (1994)) for the data fitting. Using GenSA package in R we have fitted our proposed distribution. We apply the values of log likelihood function (Log), Kolmogorov-Smirnov (K-S), P-value of (K-S) statistics to verify which distribution better fits these data. The model selection was carried out using the AIC (Akaike information criterion), the Second Order of Akaike Information Criterion (AICc) and the BIC (Bayesian information criterion)

$$AIC = -2 * l(\hat{\Theta}) + 2k$$
$$AICc = AIC + \frac{2k(k+1)}{n-k-1}$$
$$BIC = -2 * l(\hat{\Theta}) + k * Log(n)$$

where,($\hat{\Theta}$) denotes the log-likelihood function evaluated at the maximum likelihood estimates, (k) is the number of parameters, and (n) is the sample size.

We compare the results of our distribution TSN(μ,σ,λ), with the corresponding distribution of the skew-normal SN(μ,σ,λ) distribution Azzalini (1985),Skew logistic SL(μ,σ,λ) ,Alpha Skew-Normal Distribution $ASN(\mu,\sigma,\alpha)$ and normal N(μ,σ^2) distribution, MLEs, AIC, BIC, and AICc for the parameters of distributions are given in Table 1.

Table 1: MLE's, , AIC, BIC and AICc for the real data set.

Model	MLE estimate				Statistics		
	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\lambda}$	α	AIC	BIC	AICc
$\boldsymbol{N(\mu,\sigma^2)}$	7.109	1.796	-	-	813.838	820.455	724.245
$\boldsymbol{SN(\mu,\sigma,\lambda)}$	5.105	2.691	1.729	-	798.322	808.247	784.587
$\boldsymbol{SL(\mu,\sigma,\lambda)}$	5.319	2.544	1.672	-	769.062	877.998	655.067
$\boldsymbol{ASN(\mu,\sigma,\alpha)}$	6.224	1.872	-	1.542	753.012	817.145	724.019
$\boldsymbol{TSN(\mu,\sigma,\lambda)}$	4.83	1.51	2.542	-	711.70	798.06	640.06

Remark: The observed variance-covariance matrix of the MLEs of the parameters

$\widehat{\Theta} = (\hat{\mu},\ \hat{\sigma},\ \hat{\lambda})$ of $TSN(\mu,\ \sigma,\ \lambda)$,distribution from data set are

$$I_n(\widehat{\Theta}) = \begin{bmatrix} 0.1363 & 0.0337 & 0.0347 \\ 0.0328 & 0.0199 & 0.00194 \\ 0.0338 & 0.00194 & 0.0176 \end{bmatrix}$$

Table 2, provides the values of log-likelihood function (Log), Kolmogorov-Smirnov (K-S), P-value. It is evident

Table 2. The statistics -LOG, K-S, P-value for the real data set.

Model	Statistics		
	Log L	K-S	P-value
$\boldsymbol{N(\mu,\sigma^2)}$	-404.919	0.094	0.606
$\boldsymbol{SN(\mu,\sigma,\lambda)}$	-396.161	0.084	0.754
$\boldsymbol{SL(\mu,\sigma,\lambda)}$	-370.531	0.097	0.771
$\boldsymbol{ASN(\mu,\sigma,\alpha)}$	-399.452	0.095	0.740
$\boldsymbol{TSN(\mu,\sigma,\lambda)}$	-385.369	0.073	0.652

From Table 1, Table 2 that, the $TSN(\mu,\sigma,\lambda)$, distribution has the lowest statistics among all fitted models. Hence, this distribution can be chosen as the best model for fitting this data set.

Plots of the estimated density function of the real data set in Figure 4(a):

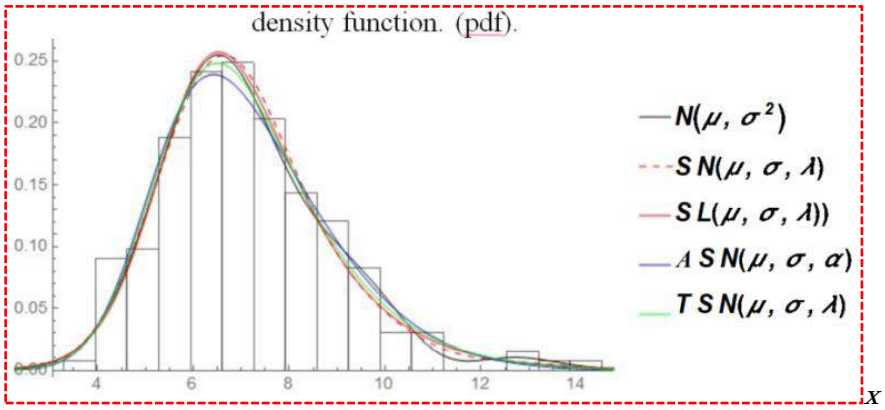


Figure 4(a): the plots of observed and expected densities of some distributions for white cells count (WCC) of 202 Australian athletes.

plots of empirical distribution and estimated cdf for the real data set.

F(x) Cumulative distributions (cdf).

© 2020 IJSRST | Print ISSN: 2395-6011 | Online ISSN: 2395-602X | Published : 18/12/2020-31-44

Figure 4(b): the plots of empirical distribution and estimated cdf white cells count (WCC) of 202 Australian athletes.

III.CONCLUSION

A new skew TSN(μ,σ,λ),distribution is constructed using a skew function (which is not a cdf) and some of its distributional properties are studied. The distribution is fitted to real life data sets and is found to perform better in real life data modeling in comparison to the skew-normal SN(μ,σ,λ) distribution Azzalini (1985),Skew logistic SL(μ,σ,λ), Alpha Skew-Normal Distribution $ASN(\mu,\sigma,\alpha)$ and normal N(μ,σ^2)equally well in the other case. Investigation of more skew TSN (μ,σ,λ), distribution considering different skew functions is currently under consideration.

IV.ACKNOWLEDGEMENT

The authors would like to thank the anonymous referees for their comments. The authors declare no conflicts of interest regarding the publication of this paper.

V. REFERENCES

[1]. Arellano-Valle, R. B., G´omez, H. W. and Quintana, F. A.(2004) A new class of skew-normal distributions. Communications in Statistics :Theory and Methods, 33, 1465—1480,

[2]. Azzalini, A. (1985). Further results on a class of distributions which includes the normal ones. Scandinavian Journal of Statistics,12: 171-178.

[3]. Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution, Journal of the Royal Statistical Society, 65, 367–389.

[4]. Chakraborty S, Hazarika PJ, Ali MM (2015) A multimodal skewed extension of normal distribution: its properties and applications. Statistics, 49(4):859-877.

[5]. Chakraborty, S., Hazarika, P. J., and Ali, M. M. (2012). A New Skew Logistic Distribution and its Properties. Pak. J. Statist, 28(4), 513-524.

[6]. Cook, R.D. and Weisberg, S. (1994). An Introduction to Regression Analysis. John Wiley & Sons, New York.

[7]. Elal-Olivero, D. (2010). Alpha-skew-normal distribution. Proyecciones (Antofagasta), 29(3), 224-240.

[8]. G´omez, H. W., Venegas, O. and Bolfarine H. (2007).Skew-symmetric distributions generated by the distribution function of the normal distribution. Environmetrics, 18,. 395—407

[9]. Gradshteyn, I.S., Ryzhik, I.M.(2004). Table of Integrals, Series, and Products, 6th edn. Academic Press, San Diego.

[10]. Gupta, R.D. and Kundu, D. (2010). Generalized logistic distributions. J. Appl. Statist. Sci., 18(1), 51-66.

[11]. Hazarika PJ, Shah S, Chakraborty S (2019).Balakrishnan Alpha Skew Normal Distribution: Properties and Applications. arXiv preprint arXiv: 1906.07424.[math.ST].

[12]. Huang, W. J., and Chen, Y. H. (2007). Generalized skew-Cauchy distribution. Statistics and Probability Letters, 77, 1137-1147.

[13]. Huang, W. J., and Chen, Y. H. (2007). Generalized skew-Cauchy distribution. Statistics and Probability Letters, 77(11), 1137-1147.

[14]. Johnson, N.L., Kotz, S., Balakrishnan, N.(1995). Continuous Univariate Distributions, vol. 2. Wiley, New York :https://onlinelibrary.wiley.com/

[15]. Louzada, F., Ara, A., and Fernandes, G. (2017). The bivariate alpha-skew-normal distribution. Communications in Statistics-Theory and Methods, 46, 7147-7156.

[16]. Ma, Y. and Genton, M. G. (2004). A flexible class of skew symmetric distribution. Scandinavian Journal of Statistics 31, 259–468. MR2087837

[17]. Mahmoud.R.M, El-Sherpieny,A, Ahmed.M.A (2015)TheNewKumaraswamyKumaraswamy Family of Generalized Distributions with Application, Pakistan Journal of Statistics and Operation Research.11, 159-180.

[18]. Mudholkar G.S.. Hutson, A.D. (2000). The epsilon-skew-normal distribution for analyzing near-normal data. J.Statist.Plan.Inference.(83) 291-309.

[19]. Prudnikov, A.P., Brychkov, Y.A., Marichev, O.I.: Integrals and Series, vols.(1–3). Gordon and Breach Science, Amsterdam (1986)

[20]. Sharafi, M., Sajjadnia,Z.,andBehboodian, J. (2017). A new generalization of alpha skew-normal distribution. Communications in Statistics-Theory and Methods, 46(12), 6098-6111.\

[21]. Teamah, A. A. and Hanan H. El-Damrawy (2007). Random sum of Lagrange random variables, 32nd International Conference of Statistics, Computer Science and its Application, Egypt, 170-178.

[22]. Teamah,A.A, Hanan H. El-Damrawy, Shaimaa M. T. Swan. (2020). Generalized Weighted Exponential-Gompertez Distribution. Applied Mathematics,11,97-118

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

A Survey of Data Exfiltration Prevention Techniques

Peter S. Nyakomitta[1*], Silvance O. Abeka[2]

[1*]Faculty of Biological and Physical Sciences, Tom Mboya University College, Homa Bay, Kenya

[2]School of Informatics and Innovative Systems (SIIS), Jaramogi Oginga Odinga University of Science and Technology, Bondo - Kenya.

ABSTRACT

Data exfiltration is a serious cybercrime facing many organizations worldwide. Over the past few years, notable organizations such as the Google, Yahoo, the Pentagon, Iran nuclear facility and the United States military contractors and banks have fallen victims of data exfiltration. The current techniques for averting these threats revolve around firewalls, intrusion detection systems, intrusion prevention techniques, firewalls, anti-virus an anti-malware. However, despite heavy deployment of these devices, attackers still continue to wreck havoc on organizations and individuals, stealing their sensitive data. The aim of this paper was therefore to explore how the current techniques for data loss prevention fail. The results of this analysis revealed that these techniques either use whitelists, blacklists, signature-based scanning, behavioral analysis of programs which are not sufficient to counter attacks based on zero day vulnerabilities. Based on these shortcomings, a novel data exfiltration prevention algorithm is proposed towards the end of this paper. This algorithm is suggested to employ real-time traffic entropy coupled with heuristically computed functional correlations to detect data exfiltrations. The premises of this algorithm and its operations are discussed at the last section of this paper.

Keywords : Data exfiltration, algorithm, IDS, IPS, anti-virus, anti-malware

I. INTRODUCTION

Data exfiltration is a form of illegal leakage of sensitive data from a particular organizational or individual system. According to Murtaza and Naveed (2016), this caliber of intrusion is hard to catch due to its careful planning and execution, often involving insider entities to leak facilitating information such as usernames and passwords. This is normally achieved through social engineering via email attachments or links which once opened or clicked, either directs the naïve user to a malicious website which installs the malware on the system, or directs runs and executes malicious activities. In most cases, the insider entity could be a person working in the target organization or a malicious hardware component bought from an unreliable third party.

A typical scenario is where an employee in an organization plugs in a universal serial bus (USB) stick infected with a malware to a machine connected to the company intranet. This malware rides on the autorun feature and executes itself as a background process (Neeshu and Shitanshu, 2016). Afterwards, this malware gains root privileges through the exploitation of zero-day vulnerabilities inherent in software already installed on the target machine such as Acrobat

Reader and Internet explorer. Figure 1 that follows gives a depiction of the data exfiltration process.

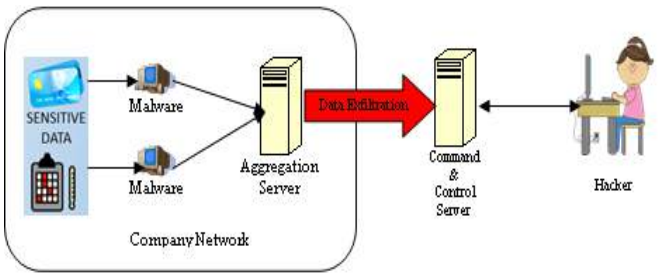


Figure 1 : Typical Data Exfiltration

As this figure shows, the company network contains sensitive data such as credit card numbers which needs to be protected from the hackers. However, an hacker can employ social engineering tactics and send the malware in form of mail attachment which upon opening, installs itself on the victim network devices. The malware then establishes a communication with the hacker controlled command and control server for additional instructions. The hacker then locates the sensitive data within the network, aggregates it together and compresses it and finally transfers it outside the organization.

Due to its stealth nature, this malware cannot be detected by the anti-virus, anti-malware or host-based intrusion detection systems since it hides behind a legitimate process through code injection. A number of these malware have modules for communicating with their designers, who can then use the obtained root privileges to install extra payload on the target machine to enable them scan the victim's machine, search and discover the exact information being sought and relay it back to attacker server before destroying itself, hence doing ways with all traces of any attack.

This makes forensic analysis of the attack a complicated affair since these malware have not only the ability to change log files but also hiding themselves behind legitimate programs. Another scenario would be for the attackers to exfiltrate data using insider hardware such as network interface cards (NICs) bought from unreliable vendors (Hanu and Dharani, 2015). The vendors might be working with attackers and hence could implant malicious firmware on the NICs such that all packets travelling from the target organization are duplicated and sent to a remote machine controlled by the attacker.

As demonstrated above, data exfiltration attacks are in most cases very difficult to detect due to their multifaceted approach, using people or hardware within organizations, malware that hides and destroys themselves after a given period of time, their ability to masquerade as legitimate programs to prevent detection, and the malware ability to communicate with attackers to receive additional payloads or instructions. In the following section, a number of techniques that are currently being employed to curb these attacks are presented.

II. CHALLENGES OF CURRENT DATA EXFILTRATION PREVENTION TECHNIQUES

Due to their ability to completely bring down organizations by disclosing sensitive and classified information, data exfiltration has received much attention from both the academic field and the market field. A number of techniques and devices have been developed to try and at least detect early phases of data exfiltration. According to Barbara (2013), firewalls employ policy based approach to detect and prevent attacks and can therefore be utilized to deter data exfiltration at the reconnaissance phase during which port scans and internet protocol IP) sweeps are carried out to identify potential targets and vulnerabilities. Figure 2 demonstrates how a firewall works.

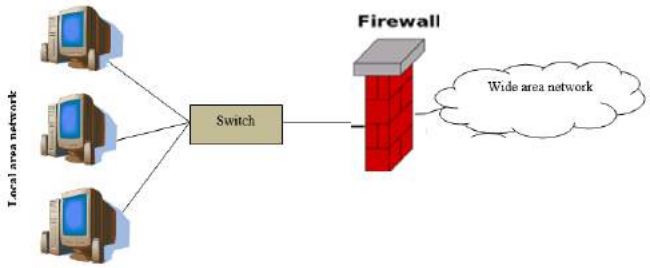


Figure 2: Firewall Operations

This figure illustrates that the firewall sits between enterprise local area network and the public wide area network such as the internet. It filters all the traffic coming into the local area network as well as that leaving the local area network. However, as Sara et al., (2016) points out, even with the presence of firewalls, the malware can encrypt and route messages and information through allowed routes such as the hyper-text transfer protocol's port 80 to the command and control servers. This is achieved through tunneling in which malicious payload is hidden in legitimate traffic. The second device that can be utilized to curb data exfiltration is an anti-virus. Randy (2015) explains that although these anti-viruses may or may not be able to contain malicious code, they can prevent a number of Trojans and other infections from spreading to other network devices by examining the files, folders, mail messages and web pages in the computers. When the scanner detects something that looks like a virus, it quarantines the suspect object and warns you about what it found. Figure 3 provides an illustration on the operation of an anti-virus.

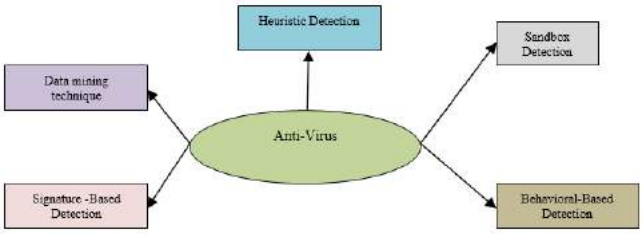


Figure 3: Operation of Anti-virus

In this figure, the anti-virus is shown to utilize five detection mechanisms for malicious software, namely the data mining, signature-based, heuristsic, behavioural and sandboxing techniques. Data mining techniques as employed in intrusion detection fall into one of two categories which are misuse detection and anomaly detection (Nadiammai and Hemalatha, 2013). In misuse detection, each instance in a data set is labeled as normal or intrusion and a learning algorithm is trained over the labeled data. These techniques have the ability of automatically retraining intrusion detection models on different input data that include new types of attacks, as long as they have been labeled appropriately. The strength of misuse detection is that it has a high degree of accuracy in identifying known attacks and their variations. Their main challenge is their inability to detect attacks whose instances have not yet been observed. On the other hand, anomaly detection builds models of normal behavior, and automatically detects any deviation from it, flagging the latter as suspect. The drawback of anomaly detection is its high rate of false alarms due to previously unseen yet legitimate system behaviors that may also be recognized as anomalies, and hence flagged as potential intrusions.

An antivirus using heuristic detection seeks to detect malware by examining the code in a malicious program and analyzing its structure. It might run a process that simulates the actual running of the code it is examining. This is meant to identify additional code logic that may help it determine if the suspected virus is really a threat. A behavioral detection antivirus program works by looking for odd behavior in a program. It monitors the operating system, searching for suspicious events. As an illustration, if the antivirus program notices an attempt to change or modify a file or communicate over the web, it may take action and warn the user about the threat. It may also block the threat depending on its security settings. Sandboxing is

normally employed to separate running programs, and executing untested or un-trusted programs or codes from unverified third parties, suppliers, users or websites, so as to mitigate any risk the malicious program may pose to the host machine or operating system.

On the flip side, Nikolaos (2015) point out that malware can hide in locations not scanned by anti-virus software. Consequently, during scans, the anti-virus may give erroneous reports regarding malware activities. In addition, some malware have the ability to stop their activities for some time so as to prevent their detection.

On their part, Intrusion Detection system (IDS), network intrusion detection systems (NIDS) and intrusion Prevention Systems (IPS) either use signature-based or anomaly based approaches. For the former case, all malicious pieces of software signatures are contained in an IDS or IPS database and the detection process involves comparing the signature of a given piece of software against the database signatures (Hanu and Dharani, 2015). As such, malicious software is identified when its signature matches one or more signatures in the database. Figure 4 is an illustration of the operation of an IDS.

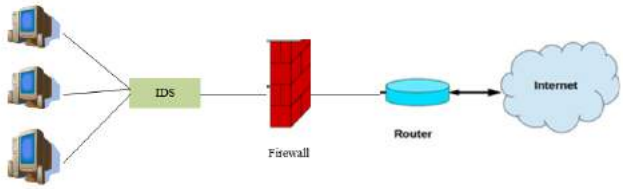


Figure 4: Intrusion Detection System

This figure points to the fact that an IDS sits between the intranet and the organizational firewall. It scans and detects any malicious activities within the network. However, as Barbara (2013) discusses, this approach is similar to the blacklist approach and can therefore not detect attacks for which a signature has not yet been created, such as zero-day exploits. Figure 5 provides a pictorial view of an IPS.

Just as was the case with IDS, the IPS sits next to the local area network to detect any anomalous activities. In addition to active detection, an IPS can actually carry out some preventive measures and therefore requires some form of manager to dictate the actions to be taken once malicious activities have been flagged down.

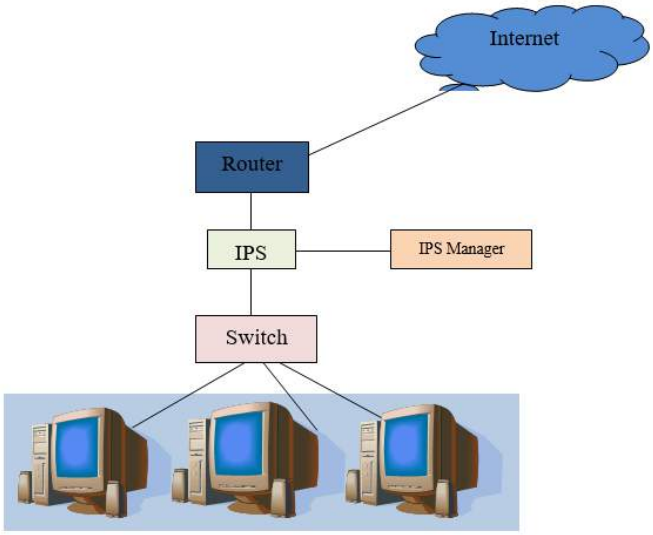


Figure 5: Intrusion Prevention System

The anomaly based approach relies on behavior analysis using machine learning to detect malicious piece of code in computer or network systems. Here, the behavior is classified as being acceptable or unacceptable and requires that a number of data sets be employed to train the learning agent o what constitute normal and abnormal behavious. As such, is a program's behavior deviates from what is considered normal behavior, it is flagged as being malicious.

However, machine-learning techniques are ideal in finding events similar to ones observed previously (Kateryna, 2017).Consequently, IDPS and IPS approaches demonstrate promising detection efficiency for specific training data sets, but are faced

with grave operational limitations when used in operational environments. In addition, behavioral analysis may fail to detect known attacks that could easily be detected with signature-based IDS if these attacks do not differ significantly from what the system establishes to be normal behavior. Figure 6 provides an overview of the operation of the machine learning malware detection process.

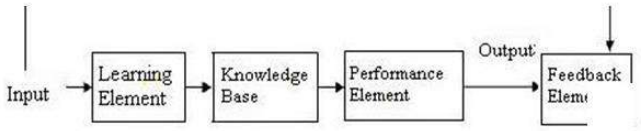


Figure 6: Machine Learning For Malware Detection

From this figure, it is clear that machine learning requires an input, which is the malicious code or program to be analyzed. It also has a database of all normal or abnormal behaviours and a learning element to detect the malware behavior. The performance element is crucial for providing feedback on the detection ability of the learning element.

Another technique that has gained popularity in as far as data loss exfiltration is concerned is the usage of anti-malware that make use of signature based scanning while a few of them use anomaly based approach. On the flip side, this anti-malware faces the same challenges as IDS and IPS. Ruchika (2013) points out that a grave concern is that attackers may test a great number of antivirus or anti-malware products and adapt their malware accordingly to escape detection.

It is clear that the current data exfiltration prevention devices face serious challenges in detecting the initial stages of the data exfiltration processes which normally involve the usage of malware to identify vulnerabilities that can be exploited to leak data from the system. If these initial stages of malware infiltration are not detected and mitigated, the then malware reports back to the attacker who can advance to gaining administrative privileges and therefore access to sensitive data. Once this is accomplished, the data is amalgamated, compresses, encrypted and tunneled through HTTP or HTTPS in the full presence of the firewall.

III. STATE OF THE ART DATA EXFILTRATION MITIGATION APPROACHES

In the wake of numerous data exfiltration attacks such as those ones involving Google, Yahoo and the U.S military, a number of researchers have come up with other techniques, methods and procedures that they believe could help mitigate data exfiltrations. One of these techniques is full packet capture which can intercept network traffic and archive it for analysis. This is ideal for after-attack forensic analysis and offer opportunity for in-depth analysis. However, according to Puneet (2013), this approach is very expensive, requires usage of external tools for low level traffic inspection, are fairly incompatible with other systems such as IDS and IPS, and provide massive data that presents storage constraints especially for high speed networks.

Another recent development in the security sector is the use of security incident and event management systems that work by collecting network events from sources such as IDS, NIDS, antivirus and event logs from firewalls. Afterwards, statistical correlation is applied on the collected events to identify probable threats. However, as Gustav (2016) explain, this technique has limited efficiency in the detection of sophisticated attacks, and has limited time window during which the event correlations and hence incidents spread over a larger time period will never be correlated. As such, a carefully executed attack in form of a series of seemingly unrelated episodes can never be detected. In addition, the event correlation is

carried out centrally and hence limited by the availability of resources.

Due to the complex and stealth nature of data exfiltration attacks, researchers are also exploring on mechanism that can help them the intruder attack vectors. Deception techniques such as honeypots, honeynets, honeywords, honeyusers and honeyfiles have been developed for this purpose. Honeypots and honeynets are deployed in internet facing and intranets to detect botnets and other attacks over wireless networks. On their part, honeyfiles have been employed to detect unauthorizes access to resources while honeywords and honeyusers have been utilized to track down compromised credentials. On the other hand, as Nikolaos (2015) explains, all these deception techniques fail to take into consideration the complicated maneuvers employed by experienced and skilled attackers. For instance, honeyfiles consist of codes that execute when the document is opened and conveys the report to the monitoring unit. An experienced attacker will decide not to open the file but exfiltrate and open it from an offline machine where the monitoring unit cannot be accessed.

IV. Proposed Algorithm For Data Exfiltration Prevention

A review of the current data exfiltration mechanisms have revealed that all of them fall short of the expectations in one way or the other. As such, this paper proposes a new data exfiltration algorithm for effective and efficient data loss prevention. This algorithm is based on information entropy, heuristics and functional correlations to detect and perform some preventive data exfiltration measures.

Information entropy will be employed to segregate plaintext and encrypted traffic. The plaintext information utilizes the twenty six letters of the alphabet and hence will have lower entropy compared with encrypted information which can make use of the alphabet, numeric and special characters. Consequently, comparing the same set of data, high entropy will be observed if these data items are encrypted and lower entropy will be recorded when this data items are just in plaintext.

Heuristics scanning will be employed to observe the behavior of the traffic that will be passing across the network. This behavior will be of two types: normal behavior and anomalous behavior. In case any anomalous behavior is observed, the algorithm will be designed in such a way that it will first alert and then take some preventive actions such as terminating the connection between the hacker controlled command and control centers and the victim's computer systems. On the other hand, the traffic with expected behavior will be allowed to traverse the network.

The last component of the proposed algorithm is the functional correlations of the network traffic, based on four decision trees: expected encrypted received encrypted; expected encrypted received plaintext; expected plaintext received plaintext; and expected plaintext received encrypted as shown in Figure 7. This algorithm will intercept traffic that is being passed across the network, and determine the port that is being used for communication purpose. This will inform the action to be taken based on the expected and the actual format of traffic received.

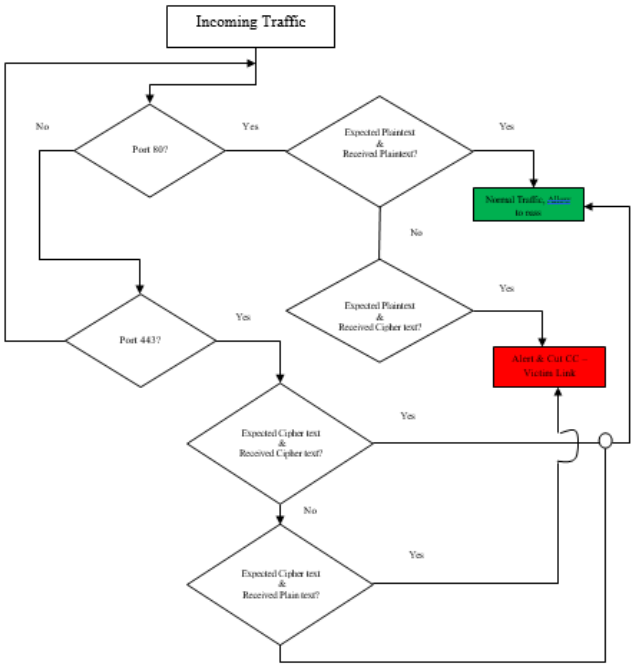


Figure 7: Proposed Data Exfiltration Prevention Algorithm Data Flow Diagram

As this figure shows, the proposed algorithm will intercept traffic, and check the port in which the data packets are being sent over. Since data exfiltration is likely to happen over hypertext transfer protocol (HTTP) or secure hypertext transfer protocol (HTTPS), the format of the data being transferred over these ports will be of great interest. In normal circumstances, the information being transferred over HTTP is in plaintext and that transferred over HTTPS is in cipher-text. Therefore, the algorithm will check to determine if this rule is being adhered to. In case of any deviation from this rule, an alert will be triggered and preventive measures such as the disconnection of the victim machine from the hacker controlled command and control (CC) will be carried out.

In so doing, the proposed algorithm will incorporate the security rules in its operation, akin to the firewall. It will also detect anomalies, just like an intrusion detection system and carry out some preventive measures in the same way that an intrusion prevention system does. As such, on top of the inclusion of the current malware detection mechanisms, the developed algorithm will incorporate some machine learning and functional correlations to supplement the signature-based approaches employed in the current IDS, IPS and anti-virus.

V. Conclusions and Recommendations

In this paper, a survey of the current and state of the art data exfiltration prevention methods and techniques has been carried out. From this analysis, it has been noted that all of them fall short of expectations in one way or the other, making it possible for attackers to carry out unauthorized data transfers from their victim machines in the full presence of IDS, firewalls, IPS and anti-virus programs. Towards the end of this paper, an algorithm that could potentially address some of the challenges noted in the current data exfiltration prevention methods has been provided. One of the pillars of this algorithm is information entropy computations that would help discriminate cipher text from plaintext. Secondly, the algorithm carries out some machine learning to segregate normal traffic behavior and anomalous traffic behavior. The last component of this algorithm is the functional correlation that will be bound to the communication ports, where four branches of the decision tree will emanate from and help in the termination of malicious activities. Owing to its strengths, this algorithm is therefore recommended for implementation in organizations that handle crucial and critical data for their customers and in military data centers.

VI. REFERENCES

[1]. Murtaza A., & Naveed G. (2016). Critical Analysis on Advanced Persistent Threats. International

Journal of Computer Applications. Volume 14, Issue No.13 (pp. 46-50).

[2]. Neeshu S., Shitanshu J. (2016). A Study and Review on Advanced Persistent Threats. International Journal of Innovative Research in Computer and Communication Engineering. Vol. 4, Issue 11. (pp. 19925-19932).

[3]. Hanu P., & Dharani J. (2015). Advanced Persistent Threat Detection System. International Journal of Science and Research. Volume 4, Issue 4. (pp. 1990- 1993).

[4]. Barbara H. (2013), Advanced Persistent Threats: Detection, Protection and Prevention. SOPHOS. (pp. 1-10).

[5]. Sara M., Arunesh S., Milind T.,and Pratyusa M. (2016).Data Exfiltration Detection and Prevention: Virtually Distributed POMDPs for Practically Safer Networks. Springer International Publishing.

[6]. Randy D. (2015). Data Loss Prevention. The SANS Institute. (pp. 1-30).

[7]. Nadiammai G., and Hemalatha M. (2013).Effective approach toward Intrusion Detection System using data mining techniques. Egyptian Informatics Journal. Vol. 13. (pp. 37-50).

[8]. Nikolaos V. (2015). Detecting Advanced Persistent Threats through Deception Techniques. Phd Thesis, Information Security and Critical Infrastructure Protection (INFOSEC) Laboratory. Department of Informatics. (pp. 1-174).

[9]. Frank K. (2015). Detection of APT Malware through External and Internal Network Traffic Correlation. Masters Thesis, University of Twente. (pp. 1-82).

[10]. Kateryna C. (2017). Machine Learning Methods For Malware Detection And Classification. University of Applied Sciences. (pp.1-93).

[11]. Ruchika M. (2013). Schemes for Surviving Advanced Persistent Threats. PhD Thesis, Department of Computer Science and Engineering, University at Buffalo, State University of New York. (pp. 1-158).

[12]. Gustav L. (2016). Bypassing modern sandbox Technologies. Masters Thesis, Department of Electrical and Information Technology, Faculty of Engineering, LTH, Lund University. (pp. 1-94)

[13]. Puneet S. (2013). A Multilayer Framework To Catch Data Exfiltration. Masters Thesis, Department of Computer Science and Electrical Engineering. (pp. 1-18).

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Troll Detection and Anti-Trolling Solution using Artificial Intelligence or Machine Learning

Saloni Dangre[1], Shubham Sharma[1], Swati Balyan[1] , Tanisha Jaiswal[1] , Dr. Pankaj Agarkar[2], Prof. Pooja Shinde[3]

[1]Department of Computer Engineering, Dr. D.Y.Patil School of Engineering, Lohegaon, Maharashtra India

[2]Head of Department , Department of Computer Engineering, Dr. D.Y.Patil School of Engineering, Lohegaon, Maharashtra India

[3]Professor, Department of Computer Engineering, Dr. D.Y.Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

With the increase in usage of social media platforms, bullying and trolling has burgeoned proportionately. The sole reason for this is that there is no surveilling authority on these platforms. To add to that, anonymity protects the identity of these bullies. Anyone from kids to teenagers to adults can fall prey to trolling. This paper focuses on using AI/ML algorithms to invigilate and report such bullies and further take actions depending on the severity of the threat imposed by them. We will be introducing lexical, aggression, syntactic and sentiment analyzers to examine a tweet and determine if it was meant to be a troll or not. The output of these analyzers will be then fed to classifier algorithms such as Naive Bayes algorithm, K-mean, to segregate these tweets based on their toxicity rating.

Keywords : Social Media; Antisocial Behaviour; Troll Detection

I. INTRODUCTION

In recent years social media has been adopted in various countries by the general publicand also by companies. Additionally, “being social”, in contrast to “being a troll”, has been shown to be vital for the standard of human interaction within the digital sphere; this attitude is often assessed in different ways. A troll remains private with an antisocial behavior that incites other users acting within an equivalent social network. In particular, a troll often uses an aggressive offensive language and has the aim to hamper the normal evolution of a web discussion and possibly to interrupt it. Only recently has it been possible to pay proper attention to the present problem, in order that many renowned press bodies and magazines have begun to address the difficulty and to write down articles both on the overall description of the phenomenon and on particular events that have caused a stir, favored by the increasing occurrence of behaviorsjust like the one described above. This type of behavior isn’t fully characterized and, up to now, it’s been difficult to seek an accurate description for the word “troll”, since the act of trolling is strongly subjective. The shortage of an agreed-on definition for the term “troll” has resulted in poor comprehension and in low interest for the research community. The necessity for

handlingthis problem has therefore emerged over time, alongsidestudies conducted by several universities and research centers. After removing applications which aren't strictly associatedwith the most topics taken into consideration (social sciences, computing and engineering), Scopus, as of 4 February 2020, lists 636 papers having the term "troll" within the title or abstract or as a keyword, when limiting the search to those three subject areas, 401 of which are related with the 2 latter topics, and 192 only to "computer science". Adding the keyword "detection" brings the entire right down to 51 papers, whose distribution in time shows a transparent increment after 2015.

For many people round the world social media sites are an integrated part of their lifestyle. There are many different social media sites supporting a good range of practices and interests. Social networks like Facebook and Twitter have become a source for news and a platform for political and moral debate for tons of users. Stories with different degrees of truthfulness are spread and tiny source criticism is applied by regular people also as journalists. The act of spreading disinformation on social media has developed from being caused by bored youths to being commercialized by organizations and political blocks within the sort of troll farms. A troll farm is a corporation whose sole purpose is to affect popular opinion with the means of social media. A practical implementation of a system or a software which will identify troll farms might be utilized in order to prevent them and thus avoid the spread of disinformation. Such an implementation would be interesting to the politicians, media, social networks or organizations that are targeted since it might be used to clear their names.

II. PREVIOUS WORK

Since 2004, with Orkut, Facebook and other social networks, people started sharing their opinions online with none moderation. Differences in opinions led to verbal spats, insults and slander. Trolling has become rampant with nobody to see on what people say within the virtual world. This inspired us to make an Anti-trolling system which will shield users online.
With the pleas of the general public getting louder and louder, Internet giants like Twitter, Facebook, Google etc. have come up with some solutions.

A. Facebook

Facebook has been the recentfavorite platform of paedophiles, bullies and hackers for an extended time. Presently users manually delete abusive messages. However, to tackle the emerging trend of online-trolls Facebook announced that they're performing on systems that automatically identify and delete abusive remarks. Administrators are being trained to handle trolls and given new tools to curb jibes. Children can report bullying and dubious behavior instantly to authorities. Official figures indicate that children below the permissible age of 13 use Facebook. Jim Gamble, the chief executive of the Child Exploitation and Online Protection Centre (CEOP) has been working with Facebook to combat cyberbullying.

Users who interact more with strangers and whose friend requests aren't accepted by other users are tagged. They maintain comprehensive grey lists to stop suspects from signing up using fake accounts. The reporting process helps to spot accounts having an equivalent IP address and each one of the fake accounts is removed.

Facebook is currently testing three new safety features around the world:

- One notifies users if someone is impersonating their account
- It analyses account names and profile pictures to find matches
- Another tool is for reporting nude photos and one more for photo safety

B. Twitter

Twitter has often been a landmine of abusive language but they're trying to maneuver from being the Wild West to a more civilized society.

Twitter is cracking down harder against trolls, including temporarily barring accounts used to harass other users. During a blog post, Twitter's vice chairman of engineering, Ed Ho announced more safety measures to prevent abuse on its platform. "Making Twitter a safer place is our primary focus and that we are now moving with more urgency than ever," Ho said in a post.One among the methods includes using the company's internal algorithms to spot problematic accounts and limiting certain account functions for a selected period of time. Twitter is additionally hospitable to further action if the harassment continued. Other anti-trolling tools include new filters to let users see what type of content they need to look at from certain accounts. They also allow people to "mute" tweets supported keywords, phrases or entire conversations. The announcement follows a series of measures that Twitter has undertaken to curb abusive behavior on its platform. The company said that it might get obviate potentially harassing tweets from feeds and searches. It also has blocked people who repeatedly swore at verified accounts.

Twitter created an automated Twitter account, Imposter Buster, which is programmed with an updated database of impersonator accounts, and every time one among them vitriolic tweets, he automatically replies and exposes them with pre-programmed evidence. But this wasn't a really foolproof method as trollers could create multiple accounts to evade the anti-trolling account.

III. LIMITATIONSOFCURRENTSOLUTIONFORON LINE SOCIALNETWORK:

Table 1- Comparison of existing anti-trolling systems

Name	Features	Strengths	Weakness
Facebook	Flag toxic posts, manual deletion of posts and comments	Human mind can comprehend toxicity more effectively	Human intervention required, Time consuming, Probability of biased results
Twitter	Imposter Buster account for monitoring tweets	Human mind can comprehend toxicity more effectively	Human intervention required, Time consuming, Probability of biased results
SCM4	Automatic Blocking of toxic tweets	User unperturbed by toxic posts	Free version blocks only up to 10 tweets per day
Perspective	Identifies toxic words and rates it according to its intensity	Identifies wide range of profanities	Still in testing phase

Table 2- Issues in perspective

Drawbacks	Original Phrase	Toxicity score	Modified Phrase	Toxicity score
Use of number	Asshole	0.98	Assho1e	0.36
Mis-spelling	Fat lady	0.73	Fatt lady	0.34
Use of special characters	Ass	0.95	A**	0.13
Use of Negation	You are a bitch	0.99	You are not a bitch	0.98
Strategic spacing	Cunt	0.96	C unt	0.12
Length of sentence	Men are superior to women	0.68	Men's biological superiority does not extend only to physical strength but also to mental faculties as compared to women	0.28

IV. PROPOSEDWORK

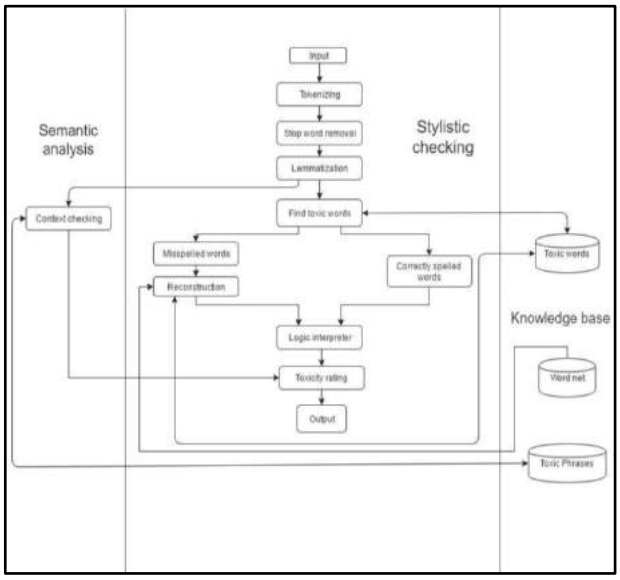


Fig. 1- Proposed Architecture

A. **Semantic Analysis:**

In this layer, the contextual meaning of the sentence is going to be analyzed.

Context checking: The precise meaning of the sentence can'tbe always understood by the literal meaning of the words utilized in the sentence. Hence during this part, the contextual meaning is taken into account.

B. **Stylistic Checking:**

Input: The tweets and other inputs are going to be accepted here in text format.

Tokenizing: The given sentence can't be easily understood by considering the whole sentence in one go. Hence, the sentence is weakened into the little part, i.e. one word per part referred to as a token. This manner helps in better understanding of the sentence.

Stop Word Removal: The words which don't contribute within the increase of toxicity of the sentence are mentioned as stop words. Such words (e.g. the, and, or) are deleted from the sentence during this step.

Lemmatization: The basic form of a word or its dictionary form is named lemma. Hence, during this part of the method, the basic form of the word is going to be returned, which can help in removing the inflectional endings and can make the method easier.

Find toxic Words: As soon as the sentence is successfully converted into tokens, the words are checked within the database for a match. (words here mean the foul, vulgar or gross language).

Misspelled Words Recognition: The words which are matched with the toxic words within the database are then forwarded to the logic interpreter after completion of indispensable reconstruction.

Correctly Spelled Words: The words which don't contain any toxicity are directly forwarded to the logic interpreted.

Logic Interpreter: After being recognized from the list and suitable reconstruction, the toxic word is passed on to the second last stage of processing where it's checked whether 'not' is employed before a toxic word. If used, as a result the comparator has got to minimize the toxicity level.

Toxicity Rating: Supported the severity of the toxicity of a word, a rating is going to be provided.

Output: The toxicity rating of the sentence as whole are going be displayed as the output.

C. **Knowledge Base:**

This domain comprises all the databases that are required for running the system. It embodies the

word net, toxic words, their ratings and toxic phrases which will be appended during semantic analysis.

V. CONCLUSION

This article has discussed the problems created by the presence of trolls in social media contexts and has presented the main approaches to their detection. In conclusion, this project has been more or less a success. In the search for trolls in the Twitter network, we managed to find our own fabricated troll farm using only the daily activity and habits of the Twitter users. We will be using the K-mean and Naive Bayes algorithm. The results showed us that both algorithms could work and could be used in the search for trolls in a social network. The K-mean and Naive Bayes algorithm is used to segregate these inputs based on their toxicity rating. The drawbacks of the current system are solved in the proposed system. The problems of spacing, special characters, negation and other drawbacks are resolved. The stylistic drawbacks as well as the contextual drawbacks of the current system are solved with the use of different analyzers as mentioned above. In this paper, we took a different approach focusing on trolling vulnerability, negative comments and algorithms to detect trolling behavior. So, by working on troll vulnerabilities and their causes we can take proactive measures to stop or minimize troll or bullying nature.

VI. REFERENCES

[1]. Zannettou, S.; Sirivianos, M.; Caulfield, T.; Stringhini, G.; De Cristofaro, E.; Blackburn, J. Disinformation warfare: Understanding state-sponsored trolls on twitter and their influence on the web. In Proceedings of the Web Conference 2019—Companion of the World Wide Web Conference, WWW 2019, San Francisco, CA, USA, 13–17 May 2019; pp. 218–226.

[2]. Badawy, A.; Lerman, K.; Ferrara, E. Who falls for online political manipulation? In Proceedings of the Web Conference 2019—Companion of the World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 162–168.

[3]. Fornacciari, P.; Mordonini, M.; Poggi, A.; Sani, L.; Tomaiuolo, M. A holistic system for troll detection on Twitter. Comput. Hum. Behav. 2018, 89, 258–268.

[4]. Donath, J.S. Identity and deception in the virtual community. In Communities in Cyberspace; Routledge: Abingdon-on-Thames, UK, 2002; pp. 37–68.

[5]. Chun, S.A.; Holowczak, R.; Dharan, K.N.; Wang, R.; Basu, S.; Geller, J. Detecting political bias trolls in Twitter data. In Proceedings of the 15th International Conference on Web Information Systems and Technologies, WEBIST 2019, Vienna, Austria, 18–20 September 2019; pp. 334–342.

[6]. "https://perspectiveapi.com/", last retrieved on 10th January 2017.

[7]. "https://jigsaw.google.com/projects/#conversation-ai", last retrieved on 14th March 2017.

[8]. "http://www.telegraph.co.uk/technology/internetsecurity/10283665/Worlds-first-anti-trolling-software-launchedin-UK.html", last retrieved on 10th October 2017.

[9]. "https://www.networkworld.com/article/2225302/opensourcesubnet/introducing-the-world-s-first--anti-trolling-software- .html" ,last retrieved on 10th October 2017.

[10]. "http://www.wired.co.uk/article/twitter-tools-harassment", last retrieved on 10th October 2017.

[11]. "http://www.dailymail.co.uk/sciencetech/article-3506068/Istrolling-Facebook-working-troll-

hunter-tool-lets-catchimpersonators-social-network.html",last retrieved on 10th October 2017.

[12]. "https://thinkprogress.org/microsofts-lovable-teen-chatbotturned-racist-troll-proves-how-badly-silicon-valley-needsdiversity-1648e7020ea6#.dwoymxuwj",last retrieved on 21st March 2017.

[13]. "http://www.tabletmag.com/scroll/219117/we-built-a-bot-thattrolls-twitters-worst-anti-semitic-trolls", last retrieved on 21st March 2017.

[14]. "World's first anti-trolling software' launched in UK", "http://www.telegraph.co.uk/technology/internetsecurity/10283665/Worlds-first-anti-trolling-software-launchedin-UK.html", last retrieved on 20th March 2017.

[15]. Phillips, Whitney; Publication Information: Cambridge, Massachusetts: The MIT Press. 2015 This Is Why We Can't Have Nice Things: Mapping the Relationship Between Online Trolling and Mainstream Culture.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Training an AI agent to play a Snake Game via Deep Reinforcement Learning

Reetej Chindarkar[1], Kartik Kaushik[1], Rutuja Vetal[1], Ronak Thusoo[1], Prof. Pallavi Shimpi[2]

[1]Department of Computer Engineering, Dr. D.Y.Patil School of Engineering, Lohegaon,Maharashtra India

[2]Assistant Professor, Department of Computer Engineering, Dr. D.Y.Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

Deep Reinforcement Learning (DRL) has become a normally adopted methodology to alter the agents to be told complex management policies in varied video games, after Deep-Mind used this technique to play Atari games. In this paper, we will develop a Deep Reinforcement Learning Model along with Deep Q-Learning Algorithm that will enable our autonomous agent to play the classical snake game. Specifically, we will employ a Deep Neural Network (DNN) trained with a variant of Q-Learning. No rules about the game are mentioned, and initially the agent is provided with no information on what it needs to do. The goal for the system is to figure out the rules and elaborate a method to maximize the score or reward.

Keywords: Deep reinforcement learning, Snake Game, Autonomous agent, Deep Learning, Experience replay

I. INTRODUCTION

One of the most well-known models of Reinforcement Learning used for playing games is called TD-gammon [1] which was developed decades ago. It was used to play the Backgammon game and it surpassed human-level performance. However, this technique shows very little generalization to different games and failed to attract wide attention. As a recent breakthrough in deep learning, DeepMind creatively combined deep learning with reinforcement learning and came up with the distinguished deep Q-learning network (DQN) model [2]. DQN outperforms all the previous approaches on six games and surpasses human-level performance on 3 games. This breakthrough lit up researchers' passion and lots of similar researches (e.g., [3, 4]) presently emerged.

However, DQN might not be straightforwardly applied to any or all scenarios as a result of its naïve reward mechanism solely produces thin and delayed rewards which will cause ineffective learning of correct policies [5]. In most reinforcement learning problems, to decrease the correlation of sampled experiences when training the network, a technique named Experience Replay is usually adopted [6]. However, this method samples previous experiences haphazardly while not considering their quality. To solve this downside, an improved approach was proposed by Schaul et al. [7] namely "Prioritized Experience Replay".

In Reinforcement Learning, we have two components: The Environment and The Agent. Every time the agent performs an action, the environment provides a reward to the agent, which might be positive or negative that depends on how smart the action was from that of the specified state. The goal of the agent is to find what actions will maximize the reward, according to the possible state. States are the observations that the agent receives at every iteration from the environment. A state may be its position, its speed, or whatever array of variables describes the environment. The Reinforcement Learning notation that is used for the decision-making method that the agent adopts is termed as policy. On a

theoretical level, a policy could be a mapping from the state space (the space of all the attainable observations that the agent will receive) into the action space (the space of all the actions the agent will take, say UP, DOWN, LEFT and RIGHT). The optimal agent is in a position to generalize over the entire state space to consistently predict the best possible action, even for those situations that the agent has never seen before.

II. GAME ENVIRONMENT AND TECHNICAL FOUNDATION

This section describes the proposed Snake Game Environment and the technical foundation required of a typical Deep Reinforcement Learning Model.

A. *Game Environment*

Snake Game is a classical digitized game, throughout which the player will control the snake to maximize the score by eating the apples that are spawned at random places. Only one apple appears within the game screen at any time. Moreover, because the snake can grow one grid long by eating an apple, avoiding collision is vital to its survival.

In this work, we will implement the Snake Game in Python as the testbed of autonomous agents. The Snake and the apple will be randomly deployed when the game starts. The game score will be initialized at 0 and will increase by 1 as the snake reaches a target. Moreover, after the collision of the snake the game will end and the game score will be reset to 0 at the start of the new game. After the previously given apple is eaten a new apple respawns and the target of the snake changes during the game when it reaches a previously determined target.

Therefore, having the ability to localize new targets in an adaptive manner is crucial for the agent playing the Snake Game. The number of control signals within the Snake Game are four, i.e. UP, DOWN, LEFT and RIGHT. At each step in time, the snake will move one grid forward on its course direction, unless the control signal it receives is in an orthogonal direction.

B. *Technical Foundation*

Deep Q-Network (DQN) was firstly presented by Mnih et al. [2] which was used to play Atari 2600 video games using the Arcade Learning Environment (ALE) [8]. DQN demonstrates its ability to successfully learn complicated control policies directly from raw pixel inputs. DQN is a convolutional neural network (CNN) trained by a variation of the classical Q-learning algorithm [9]. DQN algorithm advances traditional Reinforcement Learning techniques because it utilizes CNN to estimate the Q-function that provides a mechanism to approximate Q-values of feasible actions directly from the most recently observed states (pixels). To train the neural network and to keep the iterative evaluations stable, DQN uses mini-batches of experience. Each experience is manifested as a four-tuple (s; a; r; s'), where s denotes the state of the observed environment, a denotes the action that agent performs in state s. After the agent executes action (a) in state (s), it receives the reward r from the environment and goes into the next state (s'). Along game play, agent stores the experience in memory for future sampling and training of CNN, which is known as experience replay [6].

Additionally, DQN uses former network parameters to check the Q-values of the next state, which provides a stable training target for CNN [10]. To understand how the agent makes decisions, it's necessary to understand what a Q-Table is. A Q-table is a matrix that correlates the state of the agent with each and every potential action that the agent will take-on. The values in the table are the action's chances of success, based on the rewards it got throughout the training. The values within the Q-Table represent the expected reward of taking action a from a state s. This table is the policy of the agent that we mentioned before: it determines what actions ought to be taken from every state so as to maximize the expected reward. The problem with this can be that the policy is a table hence it can only handle a finite state space. That is to say, we cannot have an infinitely large table with infinite states. This can become a problem for those situations where we are expecting the number of possible states to be very large. Deep Q-Learning increases the capability of Q-Learning, since the policy isn't a table but a Deep Neural Network. The Q-values are updated in accordance with the Bellman equation [11].

© 2020 IJSRST | Print ISSN: 2395-6011 | Online ISSN: 2395-602X | Published : 18/12/2020 - 59-61

$$NewQ(s,a) = Q(s,a) + \alpha[R(s,a) + \gamma max Q'(s',a') - Q(s,$$

New Q-Value | Current Q-Value | Reward | Maximum predicted reward, given new state and all possible actions

Learning rate | Discount rate

Fig. 1. Bellman equation

1) *State:* A state can be the representation of a situation in which the agent will find itself. The state will also represent the input the Neural network will take.
In our case, the state can be an array which will contain 11 Boolean variables, which will take into account:
   - ❖ If the snake is in danger from its immediate proximity i.e. Left, Right and Straight.
   - ❖ If the snake moves in the direction Up, Down, Left or Right.
   - ❖ If the food is Left, Right, Above or Below.

2) *Loss:* The Deep neural network optimizes the output (action) to a specific input (state) trying to maximize the expected reward. The Loss function gives the value that expresses how accurate the prediction is compared to the truth. The job of a neural network is to reduce the loss and to abridge the difference between the real target and the predicted one.

3) *Reward:* The AI tries to maximize the expected reward in any given circumstance. A positive reward is only given to the agent once it eats the food target (+10). If the snake hits a wall or hits itself, negative reward (-10) is given. In addition, there can be a positive reward for each and every step that the snake takes without dying. In that case, the agent might just decide to run in a loop, since it would get positive rewards for each step that it takes. Reinforcement Learning agents sometimes present us with flaws in our strategy that we did not anticipate, as such outsmarting us in that way.

General Algorithm:

- ❖ The Q-value is randomly initialized at the beginning of the game.
- ❖ The system gets the present state ‘s’ (the observation).
- ❖ Based on the state 's', it executes an action, randomly or based on its neural network. During the primary phase of the training, the system usually adopts random actions to maximize exploration. Henceforth, the system depends more and more on its neural network.
- ❖ When the AI chooses and performs an action, the environment gives it a reward. After that the
- ❖ agent reaches a new state s' and it updates its Q-value as per the Bellman equation. Also, for each and every move that the agent makes, it stores the original state (s), the action (a), the state reached after performing that action (s'), the reward (r) obtained and whether the game is terminated or not. This data is then used a sample to train the neural network. This process is known as Replay Memory.
- ❖ These last two operations can be recursive in nature until an explicit condition is met (example: the game ends).

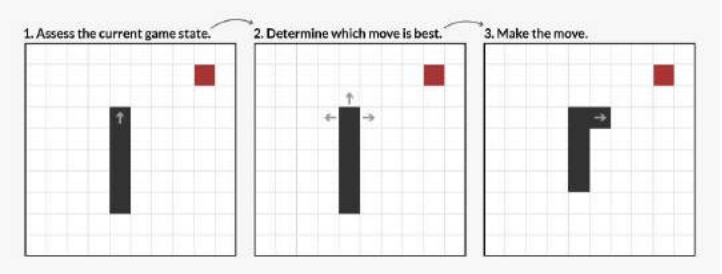


Fig. 2. Sample Image of Algorithm

III. LITERATURE SURVEY

Refer Table I for the Literature Survey

IV. CONCLUSION

In this paper, we have discussed how Reinforcement Learning can be used to train an agent that will learn how to play the classical snake game while being trained using DQN (Deep Q-Network). Specifically, we will propose a rigorously designed reward mechanism to resolve the sparse and delayed reward issue. Also, we will employ a training gap strategy to eliminate improper training experiences and implement a dual experience replay method to further improve the training efficacy. Using the State-Action-Reward-State-Action (SARSA) algorithm the proficiency of the game will increase due to the quick task performance. It will provide relatively better results as when compared to the existing techniques.

© 2020 IJSRST | Print ISSN: 2395-6011 | Online ISSN: 2395-602X | Published : 18/12/2020 - 59-61

TABLE I

Sr. No.	Paper Name	Author	Methods Proposed	Advantages	Limitations
1.	Autonomous Agents in Snake Game via Deep Reinforcement Learning	Zhepei Wei, Di Wang, Ming Zhang, Ah-Hwee Tan, Chunyan Miao, You Zhou.	Deep Q-learning	1. Using Deep Q-learning, issue of reward mechanism to solve the sparse and delayed reward issue, employ the training gap strategy to exclude improper training experiences, and implement a dual experience replay method to further improve the training efficacy has solved.	1. The first several games, both the game scores and the survival time are expected to be low because the agent only chooses random actions.
2.	Exploration of Reinforcement Learning to Play Snake Game	Ali Jaber Almalki, Pawel Wocjan	Deep Q-Learning, State-Action-Reward-State-Action (SARSA) Algorithm and Reinforcement Learning.	1. Deep Q-Learning helps to increase the efficiency of the Q-Learning and provide a steady flow to control the snake's position in the game.	1. The performance of the SARSA algorithm depends upon the user's instruction and state of the changing in the action occurred accordingly.
3.	Exploration of Reinforcement Learning to Snake	Bowei Ma, Meng Tang, Jun Zhang	Q-Learning, State-Action-Reward-State-Action (SARSA) Algorithm and Reinforcement Learning.	1. SARSA is the on-policy algorithm that helps to make decisions effectively. 2. SARSA algorithm supports agent to effectively interact with the environment.	1. Even with a decreasing exploration probability, the performance of Q-learning algorithm is not very stable. 2. Other approximation of the state space could be explored for better performance. 3. To further improve the learning rate of the Snake agent, Expected SARSA could be used.
4.	Solving the Classic Snake Game Using AI	Shubham Sharma, Saurabh Mishra, Nachiket Deodhar, Akshay Katageri, Parth Sagar	Breadth-first search, Almighty move.	1. AI Bot is trained to achieve the maximum score possible in the minimum number of steps. 2. This can also be used in other games of bigger	1. Almighty move is not used from the first iteration as the number of steps required increases to a large extent, thus increasing the time complexity. 2. In BFS, the

				size, which are the part of "Electronic Sport" to train the players.	limitation is that, it guarantees the snake till length 4, because for length greater than 4 the snake can bite itself.
5.	Automated Snake Game Solvers via AI Search Algorithms	Shu Kong, Joan Aguilar Mayans	A*Search, A* Searching with Forward, Random move.	1.The informed search algorithms can also show a reasonable reliability and the highest efficiency at the beginning of the run this property disappear at the end game. 2. In contrary almighty move is a slow algorithm at the beginning but has guaranteed a max score at end. 3. Combination of the different algorithm can achieve perfect reliability.	1. A* search algorithm is dependent on the cost of the path to reach the current fruit from the starting, and the heuristic distance from the head of the snake to the next fruit. 2. It only checks the path till the fruit is reached, with the knowledge of the previous path cost. 3. Random Move can easily reach a dead end since it blindly moves forward.
6.	Snake Played by a Deep Reinforcement Learning Agent	K. Hornik, M. Stinchcombe, H. White. Mnih et al.	Deep Reinforcement Learning, Reinforcement Learning	1. By using pixels and Convolutional Neural Networks in the state space it is possible for the agent to 'see' the whole game, instead of just nearby obstacles. 2. It can learn to recognize the places it should go to avoid enclosing and get the maximum score.	1. The agent learns to avoid obstacles directly surrounding the snake's head, but it can't see the whole game. So, the agent will enclose itself and die, especially when the snake is longer.
7.	An adaptive Strategy via Reinforcement Learning for Prisoner's Dilemma Game	Lei Xue, Changyin Sun, Donald Wunsch, Yingjiang Zhou, Fang Yu.	Reinforcement Learning, Temporal difference learning, Complex Learning.	1. The agent with adaptive strategy can make decisions under a consideration of the long-term reward. 2. The adaptive agents were able to cooperate with their opponents without losing competitiveness.	1. It is very difficult for the agents to achieve mutual cooperation by the characteristic of the scale-free network. 2. The condition of the hubs is significant for the scale free network.
8.	Deep Reinforcement	Ruben Rodriguez	DQN, Prioritized Duelling DQN and	1. It serves two advantages: it demonstrates the	1. DQNs and A2C perform badly on games with a binary

	Learning for General Video Game AI	Torrado, Julian Togelius, Jialin Liu, Diego Perez-Liebana	Advance Actor-Critic (A2C)	strengths and weaknesses of the current generation of reinforcement learning algorithms. 2. It allows results achieved on GVGAI to be compared to other existing environments.	score (win or lose, no intermediate rewards). 2. High dependency of the initial conditions which suggests that running multiple times is necessary for accurately benchmarking DQN algorithms.

V. REFERENCES

[1]. G. Tesauro, "Temporal difference learning and TDgammon," Communications of the ACM, vol. 38, no. 3, pp. 58–68, 1995.

[2]. V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," ArXiv e-prints, 2013.

[3]. E. A. O. Diallo, A. Sugiyama, and T. Sugawara, "Learning to coordinate with deep reinforcement learning in doubles pong game," in Proceedings of IEEE International Conference on Machine Learning and Applications (ICMLA), 2017, pp. 14–19.

[4]. S. Yoon and K. J. Kim, "Deep Q networks for visual fighting game AI," in Proceedings of IEEE Conference on Computational Intelligence and Games (CIG), 2017, pp. 306–308.

[5]. M. Andrychowicz, D. Crow, A. Ray, J. Schneider, R. Fong, P. Welinder, B. McGrew, J. Tobin, P. Abbeel,and W. Zaremba, "Hindsight experience replay," in Proceedings of Annual Conference on Neural Information Processing Systems, 2017, pp. 5055–5065.

[6]. L.-J. Lin, "Reinforcement learning for robots using neural networks," Ph.D. dissertation, Pittsburgh, PA, USA, 1992, UMI Order No. GAX93-22750.

[7]. T. Schaul, J. Quan, I. Antonoglou, and D. Silver, "Prioritized experience replay," Computing Research Repository, vol. abs/1511.05952, 2015.

[8]. M. G. Bellemare, Y. Naddaf, J. Veness, and M. Bowling, "The arcade learning environment: An evaluation platform for general agents," Journal of Artificial Intelligence Research, vol. 47, pp. 253–279, 2013.

[9]. P. D. Christopher J. C. H. Watkins, "Q-learning," Machine Learning, vol. 8, no. 3-4, pp. 279–292, 1992.

[10]. M. Roderick, J. MacGlashan, and S. Tellex, "Implementing the deep q-network," ArXiv e-prints, 2017.

[11]. Ali Jaber Almalki, Pawel Wocjan, "Exploration of Reinforcement Learning to Play Snake Game" in Proceedings of International Conference on Computational Science and Computational Intelligence (CSCI), 2019, pp. 377-381.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Crime Awareness and Registration System

Sudarshan Jagdale[1], Piyush Takale[1], Pranav Lonari[1], Shraddha Khandre[1], Prof. Yogesh Mali[1]
Department of Computer Engineering, Dr. D.Y.Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

The Crime rate has always been very high in India. Country has been on top in list for most of the years. The most relevant reason of such crime rate is slow judicial process and absence of knowledge in field. Many of the crimes can be prevented or reported early with an efficient user friendly system. The perspective of citizens about the judicial system and police has not been so good or falsely understood. Many of the cases do not get registered for above reasons. The reachability is poor as well as do not work for everyone. Records can be easily destroyed or fabricate. Law and rules are not known by everyone in this country and in many of the cases victim does not know that they have been accused or any criminal activity is happening around them this absence of knowledge leads to increase in crime but decrease in registering complaints. Hence crime happens which damages the society and we may not be able to overcome the situation because of absence in records and knowledge. We can overcome this gap using an efficient user friendly system which allows us to know about crimes around us and help us to file complaint of crime easily. A platform for both user and authority to connect with each other.

Keywords: Machine learning, Neural networks, IPC, Chatbot.

I. INTRODUCTION

This system has static and dynamic behaviour as most of the system the registration and filing complaint part is static where we do not need any other mathematical computation to provide these services. Where steps are predefined and data goes through the fixed number of stages and either accept or reject the request. But system also contains a Chatbot which is as important as the other static systems. Core function of system is to handle crime related queries and respond with user understandable manner. Chatbot replies with the text or with the other informational representational formats. Information which will be provided by the chatbot goes through various steps. The origin of the information is raw data which contains information about criminal law. This data is converted and stored for business.

Specific purpose. Data goes through the process to make it more useful for the system.

System will hold sensitive data of users and complaint records which drags the attraction of professional

attackers and offenders. The system should be secure and robust enough against these attackers.

II. LITERATURE REVIEW

Literature	Paper Published year	Method/description
Huang, J., Zhou, M. and Yang, D	2007	Extracting Chatbot Knowledge from Online Discussion Forums. In: 20TH International Joint Conference on Artificial Intelligence, Iijcai 2007, 423 - 428.
Fryer, L. and Carpenter, R.	2006	Bots as language learning tools. Language Learning and Technology. 10 (3) 8-14
Kumar, P. N. V.	2016	Growing cybercrimes in India: A survey. Proceedings of 2016 International Conference on Data Mining and Advanced Computing, SAPIENCE 2016, 246–251.
Nappinai N. S.	2010	The author in his paper "Cyber Crime Law in India has law kept pace with Emerging Trends? An Empirical Study" highlighted some important provision of the criminal laws in India relating to data protection, privacy, encryption and other cyber crime activities and to the extent said provisions are enforced to fight not just the present but future trends in Cyber Crime
Satow, L.	2017	Chatbots as teaching assistants: Introducing a model for learning facilitation by AI bots, 07 12Retrieved from https://blogs.sap.com/2017/07/12/ chatbots-as-teaching-assistants-introducing-a-model-for-learning-facilitation-by-ai-bots/. Schmulian, A., & Coetzee, S. A. (n.d.).
Rohit J. Kate Yuk Wah Wong Raymond J. Mooney	2005	Learning to Transform Natural to Formal Languages

Table 1. Literature Review

III. EXISTING SYSTEM

There are two ways to file our complaint in country. These types have been practiced for several years now. One of them includes physically visiting the police station or a police. Other one is registering online and communicating through the available ways at convenience.

© 2020 IJSRST | Print ISSN: 2395-6011 | Online ISSN: 2395-602X | Published : 18/12/2020- 62-72

A Visit and oral.

This is most famous and well practice type of filing complaint. In which the general process is to visit the police station or a police and record the statement by yourself if you are victim or you can file complaint on behalf of other when certain conditions match. Then police will take action on it with respect to law and procedure. After this the victim or person which performing all the activities on behalf of the victim needs to pay the visit to police station or police for several times where the police have total control over the investigation and visiting process.

This has actual human person contact which ensures the existence of a person with respect to information. This point is important in modern world of artificial intelligence, Cybercrime and all related technologies which has ability to fake one self.

Other than investigation time this process is quite time consuming and hectic for victim. In many cases victim gets harassed by the process or authority this leads to delay for justice or injustice where victim has to go through the process over and over. By statistics % of the victims founds this experience inconvenient and helpless.

We can see the loop in above diagram where victim might need to pay visits time to time for further procedures. There is also other option of phone call which is not practiced well in country due to tampering, sniffing, inaccessibility of certain resources and more importantly validation.

This plays an important role in investigation where we in most of the cases police needs to contact personally withvictim. Victim may have to go through this loop for number of times to get justice. Which is not the ideal process for all times.

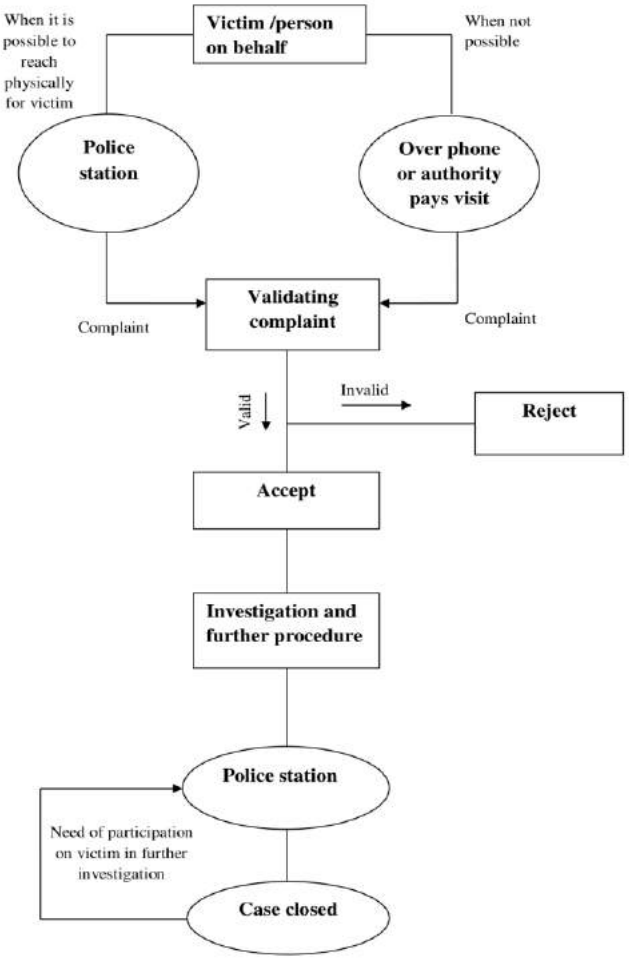


Fig. Traditional/Existing System

B.Online registration

We can observe that the first information report (FIR) only can be filed with the system other than that we have to follow same procedures again and again.

The online registration system is not consistent in country each and every state, districts and regions have different portals for same activity. Portals are also less interactive and less user-friendly which makes simple process more complicated and victim tends to choose the offline way to achieve this activity. Many of the systems only responsible for filing the very first report after that it does not interact with victim and does not hold any interaction management system for user and authority.

We observe these absence of services which can be easily filled with user-friendly and secure system we are proposing.

IV. CHATBOT

Chatbot is a conversational bot which assist you in business specific or a general manner in our project we are using this as business specific purpose where our purpose is to deliver the information about different criminal laws and acts in understandable manner. Complete chatbot system's function is to understand the scenario analyse the situation and replay with relevant output in the form of text and other information representational format if needed. Chatbot works on the predefined inputs and outputs (message and reply) and also on the dynamically generated reply. To make it more interactive and user-friendly we need to train the bot with appropriate machine learning algorithm. After that the model will able to give response in the form of business requirement which will be useful to get the actual data to be send to user. To train the model we need to feed the specific amount of data to model this data may or may not be out final output which user wants. After training process we will able to use the model in our business model as per our need.

V. PROPOSED SYSTEM

In existing system we have loop of processes for further investigation but this process can be easy with an online interaction system for user and police which with require proper authentication and validation.

Not just interaction but the system will also provide online registration system on which user can file complaint these complaints can be view on the other side by police and proper contact can be done with interaction system or any other resources if police needs to.

Other than that system has Chatbot which solves any queries regarding the criminal law it can be also used as information system as the Chatbot is an interactive and userfriendly system built with different machine learning approaches has ability to recognize the natural language with NLP and process it and give the output in natural language. These chatbots are mainly used by companies where they assist the users on website with any queries regarding the business.

We are using the same logic to complete our business purpose which is provide information about crime laws and related information about it.

A.Registration and interaction

1. When user decided to file complaint user can go the specific resources (page).

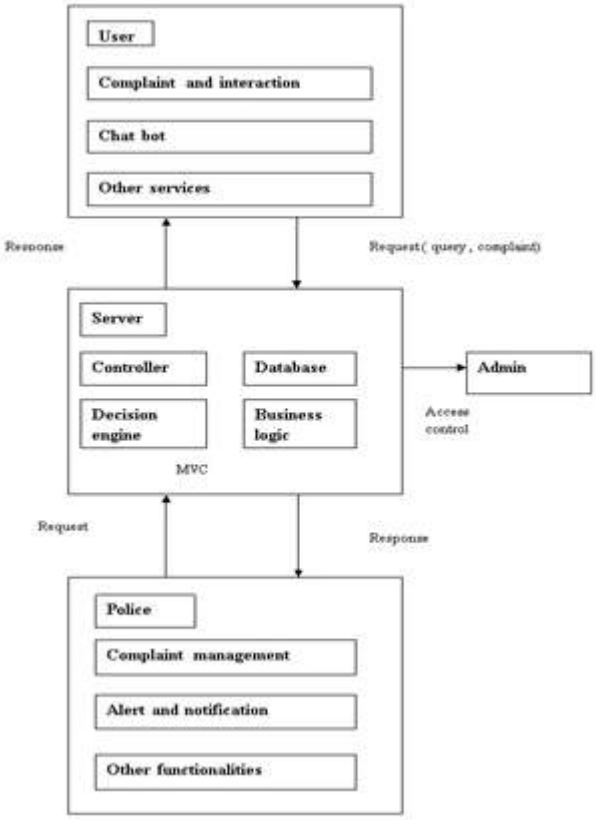


Fig. General Process

2. Enter all details about the crime happened which mainlyincludes victim, accused location, type, etc. and submit.
3. These complaint goes to the other side to police department. Now police are responsible for responding to complaint. After the verification and other procedure done police can interact with user with interaction system provided. Or contact in their own way and change the status of complaint accordingly.
4. System also provides feature to attach the media files, links and documents. Which will help police to investigate.
5. Interaction system will be robust and format as all the messages and responses will be recorded for security.

B.Chatbot

The process of using Chatbot is simple:

1. User will enter query
2. Chatbot will identify the scenario and category of information will be needed
3. Chatbot will able to give all the information about what user wants (crime related data only). Which includes the law description, statistics, precautions, next steps, any emergency helplines.
4. Based on query Chatbot will also able to recognize that if user needs to file complaint or take any actions against the crime. Chatbot will also provide the reference to registration system and act as a registration system in same interaction.

In this way user will get to know about any law which will help them for taking further actions if needed and crime rate will be less as possible.

System will not completely eliminate the traditional visit system but rather reduce it. System has a user-friendly and interactive complaint portal which will eliminate the most of the unnecessary time consuming processes. And reduces time and efforts of the victim as well as police.

Chat-bot will able to help in both scenarios as it is core purpose is to reduce human efforts and make all information available to user.

C.Other Functionalities and Services

The system can also be used to reduce the gap between the authority and citizens as well.

1. An Alert system can be used by the authority/police officers to give any alters about the crime in areas. User can be notified with alerts.
2. Missing reports of person or belongings of a person and unidentified bodies are the areas where system can play vital role. These data will be available for citizens and anyone can help in investigation using this data.
3. System can also news related to surroundings of the citizens which will make them more aware.
4. With proper personalization policies and conditions user can also help system to provide more relevant information. We can use data of user activities in system in statistics, (ex. To see most searched crimes) to understand what user needs to search which helps in surveys to understand the current standings of the country with respect to crime and criminal laws. With proper judgement and predictions this data can also help in investigation.
5. With respect to user privacy any user data will be send to the server only if user wants.

VI. RELATED WORK

Proposed system includes static as well as dynamic behaviour of system. The physical architecture is general and most used architecture where we does not need any other external device to support our requirements.

A.Physical Architecture:

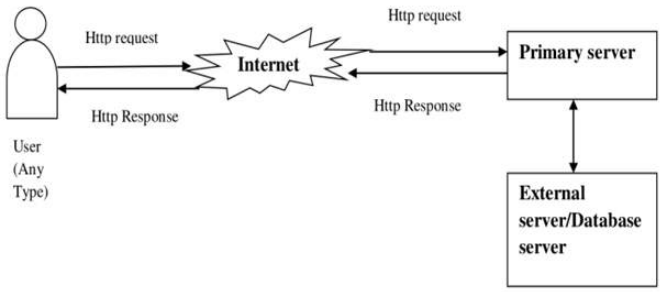


Fig. Physical Architecture

Above we can observe that this is general physical architecture

- HTTP request from user device to server using internet.
- Server (primary) process the request
- Other server may use for system database of or accessing external database (ex. Government data).
- HTTP response send from server (primary) to user after processing request with business requirements.

B.User end

- The user has various options in system
- User can chat with the Chatbot which will provide solutions to problems and answer to any question related to crime
- Helplines and other relevant services are also available on the application

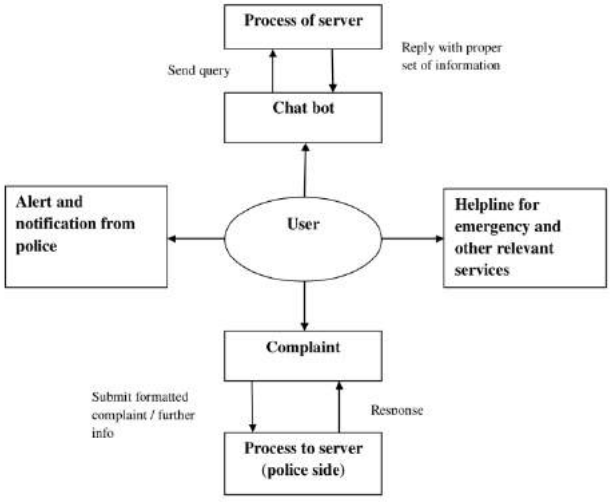


Fig. User end

- Alerts and notifications can be seen send by user
- Complaint can be register also the Chatbot will act as a complaint registration system if detects the need.

C.Police end

- Police gets complaint where they can validate the complaint, ask for more information from user and accept or reject.

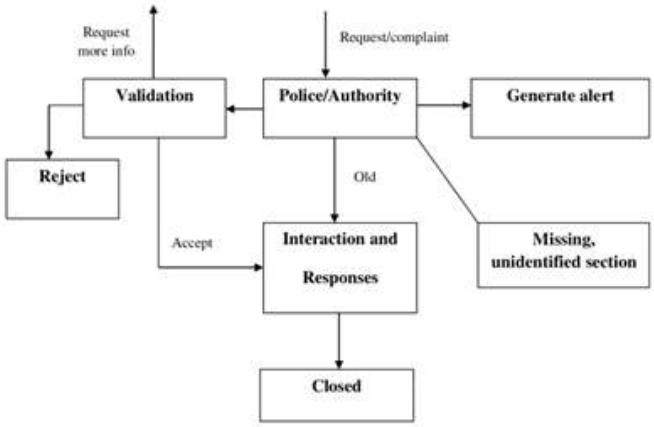


Fig. Police end

- Generate alerts in order to make citizens aware about particular activity
- Can control the missing persons and belongings section.
- Interact with user when he/she needs to.

D.Server process

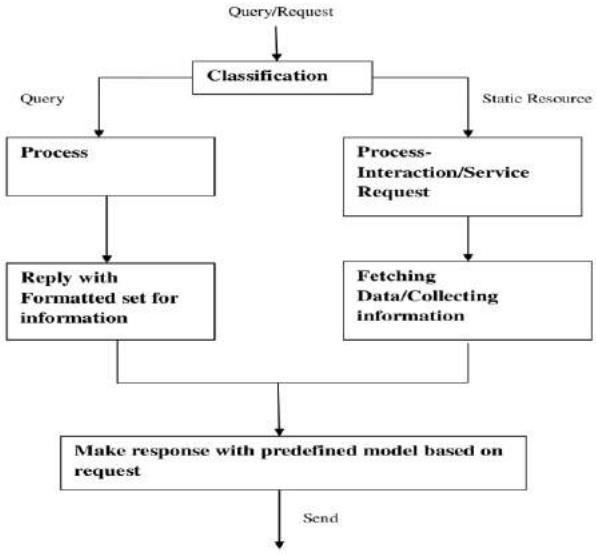


Fig. Server side

- When request gets to server it can be classified to static resources and query.
- These two requests goes from different stages.
- Static resources can be the simple database fetching and display. Where we do not need any mathematical computations to get the output.
- But in query section there are several stages. Query may go through different machine learning models to understand the natural language using different algorithms and natural language tool kits. These query then understood by machine. Which machine learning algorithms this process is done we have the numerical output processed in different stages. With the business logic this output is converted to output which is expected by user.
- These two process sends the data to response section where formatting of response is done. This response is then send to the all type of user.

E.Chatbot system

Chatbot will work as the machine/bot which chats user in natural language or as we refer the language in which human talks. To make this basic requirement fulfil first machine needs to understand the natural language because machine only understands the language of 0's and 1's other than a programming languages with set of rules.The machine cannot understand the human language. Natural language processing is the process of understanding the language with set of rules grammar. With set of mathematical rules and computations.

We implemented an algorithm for Chatbot which takes the input string and give appropriate response. We are using supervised learning model where we know the right answer and we want the model to set the rules to find right answer. This model is saved and used each time. We have to train the model for this work. For that we will use neural network. Neural networks works great in supervised learning.in neural network there is an input layer, output layer, and layers in between also called as hidden layers. While training we will decide input and output and network will make rules according to that.Our input goes through such several models, to extract category of crime, urgency and other. All have the similar training and prediction process. To make Chatbot more dynamic we have to apply more sophisticated ML algorithms which has ability to not only display the stored information but to state the same information in different forms of sentences without changing its meaning.

F. Dataset

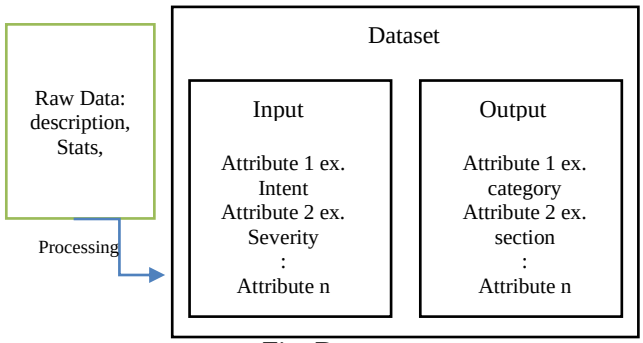


Fig. Dataset

Dataset is different than Database of the company or platform. Database contains the user information or company information. A static and mostly growing set of information. Which is used on daily basis.

On the other hand Dataset is set of information collected for training and testing the machine learning model. Machine learning model has to be trained to give expected output significant amount of times. To achieve that model has to be fed by appropriate data. We are using neural networks a deep leaning model to get our output. Input provided will be conversion of query in numerical format (neural networks only works with numbers). And output will be the array of output points in which selected neuron will hold 1 or any other number which suggest which output is chosen by model. This neural network will be using the dataset.

In this case dataset contains the dissimilar possible inputs in the form queries which can be asked by user

and the appropriate responses for those queries.This data comes from the raw data which is not necessarily in the format used by model. For that reason we have to convert the data into tables and nested objects. The data goes through number of processes to make it more usable after that it converted to multiple parameters ex. Inputs, outputs, etc. These inputs and outputs holds different parameters which also used to differentiate from one another. We can build the dataset with any flexible technology or a database which holds great flexibility of easily accessing and manipulating data. In many cases this is done by using formats like JSON, XML, xlsx, CSV or to store the data sets DB's are also used such as MongoDB, MySQL. Dataset will be divided into training and testing sets. The proportion is important. We can divide dataset in standard proportions into 80% for training and 20% for testing.

In algorithm for training and testing this dataset will be used with appropriate data structure and programming logic for conversion and feeding to model.

G.Training algorithm

1. data dataset {input, output} :
   Data is imported from the dataset where input and output is known.
2. For each input in data.input:
   a. Tokenize the input sentence into word sequence.
      Convert sentence into array of words
   b. Remove non-alpha chars. From words
      Remove all non-alphabetical characters (? ,' . :)
   c. Get base word of each word.
      Convert the words into dictionary words (doing do)
   d. Create dictionary of words from the total words in all inputs.
      Assign numbers to them the dictionary should contain only distinct words with distinct number values.
   e. Convert all words in input to numbers in dictionary.
      Neural network only works with numbers.
3. Create an output array for neural network: For all the outputs create the array of total outputs which contains only selected output as 1 and others as 0 (required for neural network).

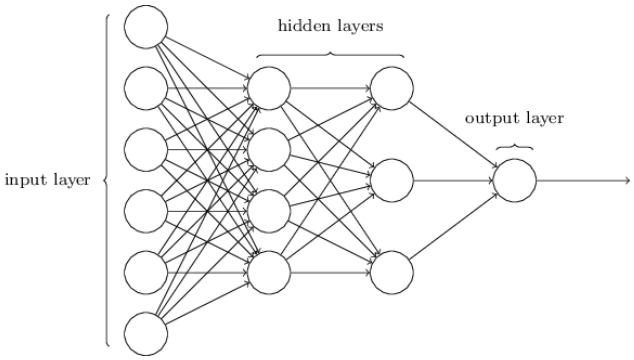


Fig. Neural Network

4. create and define a user defined model or import form library:
   a. Define number of layers.
   b. Define number of neurons for each layer. For input it is same as size of max input size. And output as number of outputs. Define number of neurons in all middle layers.

An activation function for neurons : ReLU

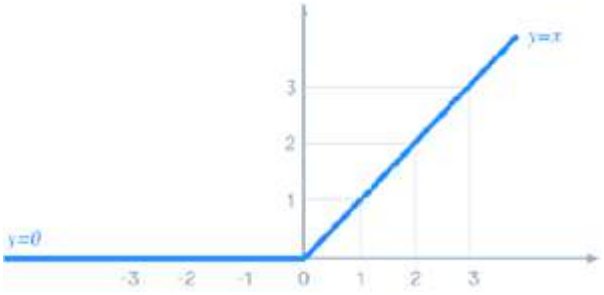


Fig. ReLU activation function

   e. Choose loss function: mean squared error chosen.
   f. Use any suitable optimizer: Adam chosen.
5. Give data to model.

6. Select epoch (number of times model has to train)
7. Start training
8. Test the model with testing set.

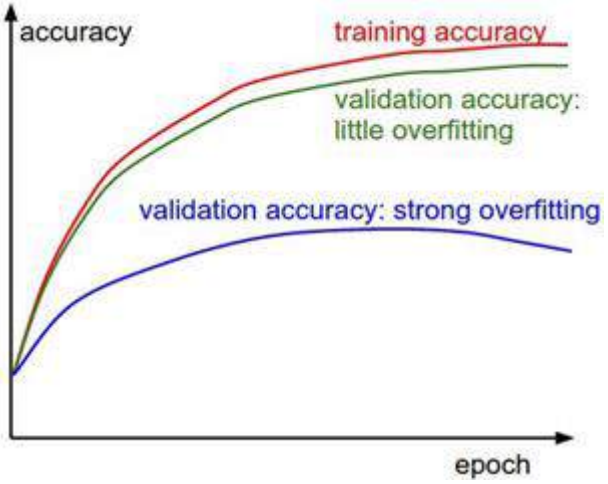
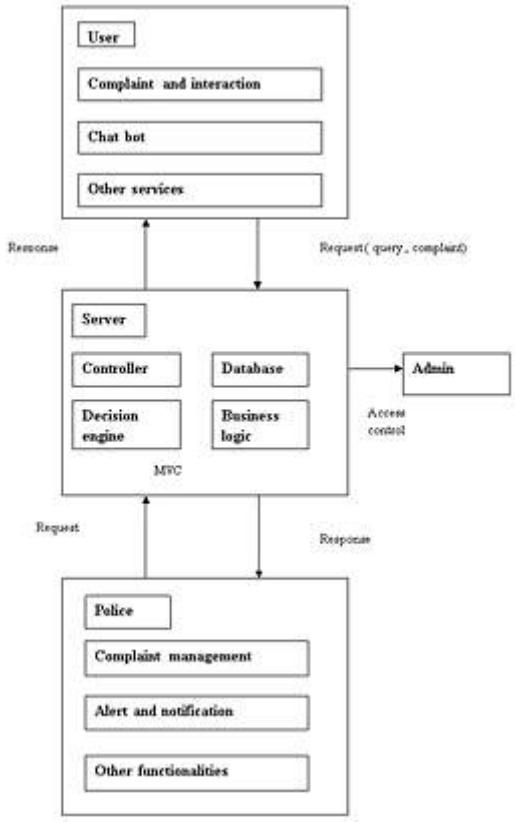


Fig. Model accuracy

9. Deploy if accuracy meets decided threshold.



H.Prediction algorithm

1. Tokenize the all words in the input. Convert array of words.
2. Convert words into dictionary equivalent numbers.
3. Give input to model.
4. Get the output (selected output neuron).

All the models in in this process works with same number of steps for different feature extractions.

I.Cross platform

System is buit with crossplaform technology where we have to put less efforts and time to develop efficient system. Which can work on multiple platfroms easily native apps is one option for that where we can Cross-platform our application with flexible layout. Native apps can easily access device resources.

VII. DRAWBACKS

- Depends on external database for user verification: As our database holds only the information collected in portal itself it depends on the government database for Adhar, PAN or any other verification.
- Needed to train models when government make amendments in law:
  In current situation of country in past 6 years government made or rather had to make amendments in the law and enforcements. With that frequency it has been observed that government will continue to do so in that case we have to retrain our model with new data.
- To use these services and get information of register complaints we need a device with working internet. In many rural areas there is an absence of such devices and inaccessibility of technology, system cannot help these citizens.
- System will require additional layers of security and extra care while controlling access: The Police or any authority type of user will have to use the system with extra care as it contains very sensitive data which attracts attention of many offenders and attackers. For that we need to train the staff for using the system.

- There is no supervisor in the system to keep eye on activities of police and other authority. We need to add more levels of user.
- Chatbot accuracy directly depend on the training model and the dataset. This idea of providing information about the crime law and related topics is not yet implemented much to draw anyone's attention hence there is lack of data and knowledge in the field.

VIII. CONCLUSION

In this paper an attempt has been made to present a new idea with current technology to solve important issues of society. With the help of natural language understanding and other machine learning algorithms we can built a system with minimum requirements and necessary services. First we discussed the traditional methods have been practiced till now. Then we discussed how we can change it with technology we have. After that we saw how system can be implemented. Taking basic needs of citizens this paper highlights on unfocussed areas of reachability of citizens towards the crime and law departments.

IX. REFERENCES

[1]. Ravichandran, K., &Arulchelvan, S. (2017). "Structural Equation Model Analyzed on Cyber Crime and Media Awareness in India". 2017 Second International Conference on Recent Trends and Challenges in Computational Models (ICRTCCM).

[2]. Zayid, E. I. M., & Farah, N. A. A. (2017). "A study on cybercrime awareness test in Saudi Arabia - Alnamas region". 2017 2nd International Conference on Anti-Cyber Crimes (ICACC)

[3]. Ki, Y. P., &Hee, Y. Y. (2011). "Crime Prevention System Based on Context-Awareness". 2011 3rd International Workshop on Intelligent Systems and Applications.

[4]. Datta, P., Panda, S. N., Tanwar, S., &Kaushal, R. K. (2020). "A Technical Review Report on Cyber Crimes in India". 2020 International Conference on Emerging Smart Computing and Informatics (ESCI).

[5]. Azeez, J., &Aravindhar, D. J. (2015). "Hybrid approach to crime prediction using deep learning". 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI).

[6]. Mkhwanazi, K., Owolawi, P. A., Mapayi, T., &Aiyetoro, G. (2020). "An automatic crime reporting and immediate response system". 2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD).

[7]. Gambino O. Augello A. Caronia A. Pilato G. Pirrone R. Gaglio S., "Virtual conversation with a real talking head. Proceedings of the Conference on" Human System Interactions, 25-27 May 2008, Kraow, Poland, pp. 263-268.

[8]. R. S. Russell, "Language Use, Personality and True Conversational Interfaces", Project Report of AI and CSUniversity of Edinburgh, Edinburgh, pp.1-80, 2002.

[9]. Y. Zhou, X. Ziyu, A. W. Black, A. I. Rudnicky, "Chatbot Evaluation and Database Expansion via Crowdsourcing", Proc. of the Chatbot Workshop of LREC, US, pp. 16-19, 2016.

[10]. C. R. Anik, C. Jacob, A. Mohanan, "A Survey on Web Based Conversational Bot Design", JETIR, Vol.3, Issue.10, pp. 96-99, 2016.

[11]. Ramos, S., Perez-Lopez, J. A., & Abreu, R. (2020). "An Analysis of the Importance of the Artificial Intelligence on the Information System of Police Forces". 2020 15th Iberian

Conference on Information Systems and Technologies (CISTI).

[12]. Mary Shermila, A., Bellarmine, A. B., & Santiago, N. (2018). "Crime Data Analysis and Prediction of Perpetrator Identity Using Machine Learning Approach". 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI).

[13]. Gonda, D. E., Luo, J., Wong, Y.-L., & Lei, C.-U. (2018). "Evaluation of Developing Educational Chatbots Based on the Seven Principles for Good Teaching". 2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE).

[14]. Murcia Trivino, J., Moreno Rodriguez, S., Diaz Lopez, D. O., & Gomez Marmol, F. (2019). "C3-Sex: a Chatbot to Chase Cyber Perverts. 2019 IEEE Intl Conf on Dependable, Autonomic and Secure Computing", Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech).

[15]. Zolfi, H., Ghorbani, H., &Ahmadzadegan, M. H. (2019). "Investigation and classification of cyber-crimes through IDS and SVM algorithm". 2019 Third International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)

[16]. Rajapakshe, C., Balasooriya, S., Dayarathna, H., Ranaweera, N., Walgampaya, N., &Pemadasa, N. (2019). "Using CNNs RNNs and Machine Learning Algorithms for Real-time Crime Prediction". 2019 International Conference on Advancements in Computing (ICAC).

[17]. Kumar, M., Hanumanthappa, M., & Kumar, T. V. S. (2017). "Use of AADHAAR biometrie database for crime investigation — Opportunity and challenges". 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS).

[18]. G. Krishna Vamsi; Akhtar Rasool; Gaurav Hajela, "Chatbot: A Deep Neural Network Based Human to Machine Conversation Model", 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT).

[19]. A. M. Rahman, Abdullah Al Mamun; Alma Islam, "Programming challenges of chatbot: Current and future prospective",2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC).

[20]. MounikaKarna; D. Sujitha Juliet; R.Catherine Joy, "Deep learning based Text Emotion Recognition for Chatbot applications", 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184).

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Emotion Recognition Based Personal Entertainment Robot Using ML & IP

Shivani Chougule, Shubham Bhosale, Vrushali Borle, Vaishnavi Chaugule, Prof. Yogesh Mali
Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

This project presents a method to automatically detect emotional duality and mixed emotional experience using Linux based system. Co-ordinates, distance and movement of tracked points were used to create features from visual input that captured facial expressions, head, face gestures and face movement. Spectral features, prosodic features were extracted using the camera. Espeak and Pyttsx and Face API was used for calculation of features. A combined feature vector was created by feature level fusion and cascade classifier was used for emotion detection. Live participants and actions are to be used for recording simultaneous mixed emotional experience. As per calculated result system will play songs and display books list.

Keywords: Smart Emotion, Espeak and Pyttsx and Face API

I. INTRODUCTION

Emotion recognition has important applications in the field of medicine, education, marketing, security and surveillance. Machines can enhance the human-computer interaction by accurately recognizing the human emotions and responding to those emotions. Existing research has mainly examined automatic detection of single emotion. But psychology and behavioral science studies have shown that humans can concurrently experience and express mixed emotions. For instance, a person can feel happy and sad at the same time. In this research combinations of six basic emotions (happiness, sadness, surprise, anger, fear, disgust and neutral state) were used. The aim of this study is to develop features that capture data from facial expressions to identify multiple emotions. In case of single-label classification problem each annotated feature-vector instance is only associated with a single class label. However, the multiple concurrent emotion recognition is a multi-label classification problem. In a multi-label problem, each feature vector instance is associated with multiple labels such as presence or absence of one of each six basic emotions. The multi-label classification is receiving increased attention and is being applied to a many domains such as text, music, images and video based systems, security and bioinformatics. This paper examined recognition of concurrent emotional ambivalence and mixed emotions. Additionally, the study examined two concurrent emotions (emotion duality) to limit the scope of the research based on availability of scenarios. This was done so that the experimental design was realistic. The subjects could express dual emotions with ease and observers could annotate the

data without ambiguity. This study implemented a multimodal emotion recognition system with multiple check box input to facilitate the annotation of concurrent emotions in the user interface software.

II. PROBLEM STATEMENT

In the recent years, Variety of emotion expression process shows that there are many ways to describe global and local speech properties, and one of the most effective is emotions. Main motivation of the system is to automatically identify users mood and according to that related books list, video list and music will play through Linux based system.

III. LITERATURE SURVEY

[1] S. Patwardhan, "Augmenting Supervised Emotion Recognition with Rule-Based Decision Model", arXiv, 2016.

Description: In this paper, we investigate the effect of transfer of emotion-rich features between source and target networks on classification accuracy and training time in a multimodal setting for vision based emotion recognition.

[2] M. Liu, R. Wang, S. Li, S. Shan, Z. Huang, and X. Chen. Combining multiple kernel methods on riemannian manifold for emotion recognition in the wild. ICMI, 2014.

Description: Emotional expressions of virtual agents are widely believed to enhance the interaction with the user by utilizing more natural means of communication. However, as a result of the current technology virtual agents are often only able to produce facial expressions to convey emotional meaning.

[3] A. S. Patwardhan, "Augmenting Supervised Emotion Recognition with Rule-Based Decision Model", arXiv, 2016.

Description: This paper presents a method to automatically detect emotional duality and mixed emotional experience using multimodal audio-visual continuous data. Co-ordinates, distance and movement of tracked points were used to create features from visual input that captured facial expressions, head, hand gestures and body movement. Spectral features, prosodic features were extracted from the audio channel.

[4] SE. Kahou, C. Pal, X. Bouthillier, P. Froumenty, C. Glehre,R. Memisevic, P. Vincent, A. Courville, Y. Bengio, RC. Ferrari and M. Mirza. Combining modality specific deep neural networks for emotion recognition in video. Proceedings of the 15th ACM on International conference on multimodal interaction, 2013.

Description: This paper presents the initial implementation of a system of multimodal recognition of emotions using mobile devices and the creation of an affective database through a mobile application. The recognizer works into a mobile educational application to identify user's emotions as they interact with the device.

[5] A. S. Patwardhan and G. M. Knapp, "Multimodal Affect Analysis for Product Feedback Assessment," IIE Annual Conference. Proceedings. Institute of Industrial Engineers-Publisher, 2013.

Description: In this paper, we investigate the effect of transfer of emotion-rich features between source and target networks on classification accuracy and training time in a multimodal setting for vision based emotion recognition.

IV. PROPOSED SYSTEM

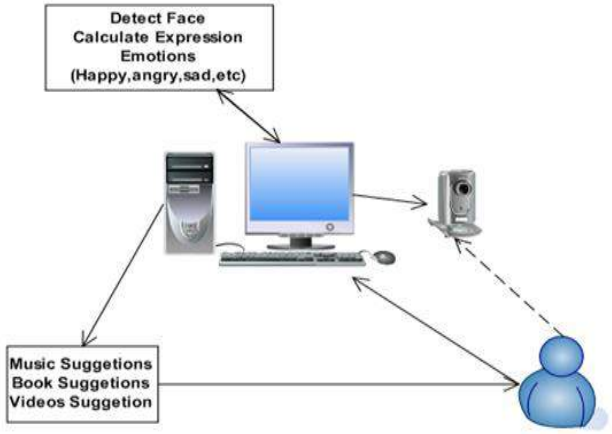


Fig 1. System architecture

User: Use this system.
Server: Connection between User and database.

Database: Storage of information related to Facial characteristics, songs and books uploaded.

Our system has mainly three modules, user module, mood detection module and video suggestion module. Various processes involved in these two modules are:

User Module:

User can use system and store songs and books library in the system.

Mood detection Module:

As per the facial expression it will recognize mood of user it will show songs list or books library and it will also give video suggestion.

Video suggestion Module:

According to the users mood it will give the suggestions of videos.

V. CONCLUSION

To conclude, music is an important means of regulating mood in various everyday situations. Proposed system is readily available to everyone and can be listened to almost anywhere. System is directly dependent upon Facial Expressions of user, so it is very effective.

VI. REFERENCES

[1]. A. S. Patwardhan, "Augmenting Supervised Emotion Recognition with Rule-Based decision Model", arXiv, 2016.

[2]. A. S. Patwardhan and G. M. Knapp, "Affect Intensity Estimation Using Multiple Modalities," Florida Artificial Intelligence Research Society Conference, May. 2014.

[3]. A. S. Patwardhan and G. M. Knapp, "Multimodal Affect Analysis for Product Feedback Assessment," IIE Annual Conference. Proceedings. Institute of Industrial Engineers-Publisher, 2013.

[4]. SE. Kahou, C. Pal, X. Bouthillier, P. Froumenty, C. Glehre,R. Memisevic, P. Vincent, A. Courville, Y. Bengio, RC. Ferrari and M. Mirza. Combining modality specific deep neural networks for emotion recognition in video. Proceedings of the 15th ACM on International conference on multimodal interaction, 2013.

[5]. F. D. Schönbrodt and J. B. Asendorpf, "The Challenge of Constructing Psychologically Believable Agents," J. Media Psychology, vol. 23, pp. 100-107, Apr. 2011.

[6]. N. C. Krämer, I. A. Iurgel, and G. Bente, "Emotion and motivation in embodied conversational agents," presented at the AISB'05 Conv., Symp. on Agents that Want and Like: Motivational and Emotional Roots of Cognition and Action, Hatfield, UK, 2005.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Automatic Whitelist Generation for SQL Queries Using Web Application Test

Venkati Mane[1], Jayesh Trivedi[1], Manalikamble[1], Shital Janjal[1], Prof. Vandana Chavan[2]

[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

[2]Professor Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

In this survey paper the proposed system will enable secure access of data to a voice- based user interface (UI) by enabling voice-based authentication and integration with an existing Natural Language Processing (NLP) system. This survey paper focuses on voice based SQL query generation. We study the question of how to improve the fetching the results from query results as well as applying the query to the database. Traditional predefined query forms are not able to satisfy various ad-hoc queries from users on those databases. Here, we propose Machine learning based technique to generate the SQL query based on user voice, a novel database query form interface, which is able to dynamically generate query forms.

Keywords: NLP, Languages and compilers, Optimization, Verification, Voice Recognisation, Machine-independent microcode generation

I. INTRODUCTION

Natural Language Processing (NLP) is an area of application and research that explores how computers can be used to understand and manipulate natural language speech or text to do useful things. The foundation of NLP lie in a number of disciplines, namely, computer and information sciences, linguistics, mathematics, electrical and electronic engineering, artificial intelligence robotics, and psychology. NLP researchers aim to gather knowledge on how hu- man beings use and manipulate natural languages to perform desired tasks so that appropriate tools and techniques can be developed. Applications of NLP include a number of fields of study such as multilingual and cross-language information retrieval (CLIR), machine transaction, natural language, text processing and summarization, user interfaces, speech recognition, artificial intelligence and expert systems.

While natural language may be the easiest system for people to learn and use, it has proved to be the hardest for a computer to understand. The goal of NLP is to enable communication between people and computers without resorting to memorization of complex commands and procedures.

In other words, NLP is a technique, which can make the computer understand the languages naturally used by humans. In this project, we are translating English query

Into a SQL query using semantic grammar. The system will accept users query in natural language as an input. The program will check whether the query is valid or not.

Then we will generate tokens by performing the division of the question clause. Each token represents a single word in the users query. The tokens from the query clause are compared with clauses already stored in the dictionary. The dictionary needs to be constantly updated. Then the algorithm scans the tokens and tries to find attributes present in the query. Then we find all the tables in the database which contain the attributes by comparing syntax and semantics. Then we build the final SQL query and execute it on the database and return the result dataset to the user.

II. PROBLEM STATEMENT

Most of used don't know about the data access, data searching, data query so, user have better knowledge that area. A voice controlled system for blind, which transceivers information in the form of audio: a personal virtual assistant which can take the human voice commands to perform tasks which otherwise would need the dependence on others.

III. LITERATURE SURVEY

Natural language processing can be done in two way communication with device one is written communication as well as verbal communication with device written communication is much more easier than the verbal communication. In written communication syntax, semantic, lexical and morphological analysis is done. Whereas in verbal communication includes all the process in written as well as additional process include additional knowledge about phonology as well as enough added information to handle the further ambiguities that arise in speech.

In this paper author represent a method for building a ldquonatural languages interfaces to data basesrdquo (NLIDB) system. [1].

system which is open source and can be accessed in remote locations using a simple hardware component. The processes including voice extraction, speech to text conversion, text processing and database management and speech synthesis have been explained in detail along with the python modules used to build the system. By minimizing the hardware components and using open source software, a universal, adaptable NLP system has been proposed [2].

The paper discusses natural language processing in field of Information retrieval. Natural Language Processing deals with human computer interaction. Natural language processing has a vital role in the field of Information retrieval. The storage and processing of information in response to user's query requires an information retrieval system that processes human or natural language. Various tasks of information retrieval such as Stemming, Lemmatization, POS tagging is discussed in the paper. The implementation of these tasks in Python is illustrated with examples.

The paper explains and implements natural language processing task in Information retrieval. Essentail and preliminary tasks task required in almost all information retrieval system is explained with example in the paper. The tasks are performed in Python and with the inclusion of NLTK module. [3].

In this paper proposed system accepts user query in natural language and translate it into SQL query and retrieve result from database. Syntactic parsing, keyword extraction, stop words removal, cooccurrence matrix generation, use of WordNet, stemming algorithm and semantic mapping techniques have been used for formation of the SQL query from natural language input. Developed system gives correct answers of simple queries, queries with logical conditions and aggregate functions. As presented system does not support all forms of SQL queries, further development is necessary.

In this paper author discussed design and implementation of a system using modified word co-occurrence matrix method which will provide access to database using queries in English language [4].

In pepar how to work are Recognizing speech using Sphinx, to voice base natura language processing using python [5]

We introduce a stochastic model for dialogue systems based on the Markov decision process. Within this framework we show that the problem of dialogue strategy design can be stated as an optimization problem, and solved by a variety of methods, including the reinforcement learning approach. The advantages of this new paradigm include objective evaluation of dialogue systems and their automatic design and adaptation. We show some preliminary results on learning a dialogue strategy for an air travel information system[6]

In this paper, we use information retrieval (IR) techniques to improve a speech recognition (ASR) system. The potential benefits include improved speed, accuracy, and scalability. Where conventional HMM-based speech recognition systems decode words directly, our IR-based system first decodes subword units. These are then mapped to a target word by the IR system. In this decoupled system, the IR serves as a lightweight, data-driven pronunciation model. Our proposed method is evaluated in the Windows Live Search for Mobile (WLS4M) task, and our best system has 12% fewer errors than a comparable HMM classifier. We show that even using an inexpensive IR weighting scheme (TF-IDF) yields a 3% relative error rate reduction while maintaining all of the advantages of the IR approach[7]

IV. PROPOSED SYSTEM

A. System Architecture:

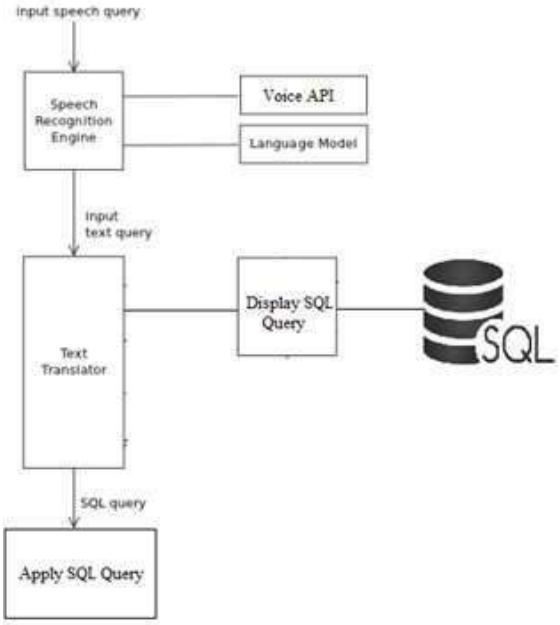


Fig 1. System architecture

This system has been under developed to enable secure access of data to a voice-based user interface (UI) by enabling voice- based authentication and integration with an existing Natural Language Processing (NLP) system.

We address the question of how to improve the fetching the results from query results.

Here, we propose Machine learning based technique to generate the SQL query based on user voice, a novel database query form interface, which is able to dynamically generate query forms.

B. Description:

1. In this system we implement the sql as a voice command and apply Natural Language Processing (NLP) with the help of Machine Learning technique.

2. First we use voice based input and after the input it is processed in the speech recognition engine.

In the speech recognition engine it has voice API and Language Model the 2

3. components of the speech recognition engine. Then the input text query is processed in the text translator and from the translator it is sent and then displayed in sql and then stored in the database that stored the sql query which will be applied in database. After the text translator query will be saved in the whitelist then the query will apply in the database.

V. APPLICATION

- This system scope is not limited. we use this system to the non-technical as well as technical user for understating the SQL Query.
- This kind of human-machine interfaces would allow a user to control a Database activity using the voice based.
- Government and private school.
- Database Management Company

VI. CONCLUSION

The implementation of the proposed system aims to translate SQL query into speech (voice). The scope of the project is to enhance the recognition capability for various SQL query and achieving more accuracy to generate the Database.

We address the question of how to improve the fetching the results from query results. Traditional predefined query forms are not able to satisfy various ad-hoc queries from users on those databases. Here, we propose Machine learning based technique to generate the SQL query based on user voice, a novel database query form interface, which is able to dynamically generate query forms.

VII. REFERENCES

[1]. F.Siasar djahantighi, M.Norouzifard, S.H.Davarpanah, M.H.Shenassa, "USING NATURAL LANGUAGE PROCESSING IN ORDER TO CREATE SQL QUERIES", in Proceedings of the International Conference on Computer and Communication Engineering, Kuala Lumpur, May 2008, pp. 600-604.

[2]. NATURAL LANGUAGE PROCESSING USING PYTHON International Journal of Scientific Engineering Research Volume 8, Issue 5, May-2017 19 ISSN 2229- 5518IEEE]

[3]. Natural Language Processing Techniques Applied in Information Retrieval-Analysis and Implementation in Python, TulikaNarang, International Journal of Innovations Advancement in Computer Science IJIACS ISSN 2347 – 8616 Volume 5, Issue 4 April 2016

[4]. Anuradha Mohite, Varunakshi Bhojane, "Natural Language Interface to Database Using Modified Co- occurrence Matrix Technique 2015 International Conference on Pervasive Computing (ICPC)

[5]. Levin E., Pieraccini R., Eckert W. "Learning dialogue strategies within the Markov decision process framework", Automatic Speech Recognition and Understanding IEEE Proceedings, pp. 72-79, 1997.IEEE]

[6]. Liddy, E. D. In Encyclopedia of Library and Information Science, 2nd Ed. Marcel Decker, Inc.

[7]. J. Allan. Perspectives on information retrieval and speech. In Information Retrieval Techniques for Speech Applications: LNCS 2273, pages 1–10, 2002

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

IOT Based Smart Electric Meter

Shikha Kushwaha[1], Sahil Dhankhar[1], Shailendra Singh[1], Vishal Kisan Borate[2]

[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

[2]Assistant Professor Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

Electricity plays a cardinal role in day to day life. The electrical energy consumption in India is the third biggest after China and USA with 5.5% global share in 2016. Due to manual work, our existing electricity billing system has major drawbacks. This system gives the information on meter reading, power cut and the alert systems for producing an alarm when energy consumption exceeds beyond the specified limit using IoT.This idea is being implemented to reduce the human dependency to collect the monthly reading and minimize the technical problems regarding billing process. From the electricity board section, the information regarding the bill amount, payment and the pre-planned power shut down details are communicated to the consumer. If the customer does not pay the bill in time, the user is informed through a message. In the already existing smart energy meter, it shows the energy consumed by the appliances from the date of installation of the energy meter and its corresponding rupees. In this proposed energy meter, the meter gives the energy consumed on daily basis, its corresponding rupees, billing details and payment using IoT. This system not only reduces the power cut issues and the labor cost for noticing the residential energy consumption in regular intervals but also increases the energy conservation with the help of alarm systems and the energy meter accuracy by reducing the billing error and the cost of maintenance.

Keywords—Arduino, GSM, IoT, energy consumption, human dependency, shut down, alert message, payment details, daily basis, alarm systems.

1. INTRODUCTION

Monitoring and keeping tracking of electricity consumption for verification is a tedious task today since manual meter reading and recording is in vogue. It is important to know from the customer view point that if one is charged fairly and according to the need. [2]

Automation of the system will allow users to monitor energy meter readings over the internet in the real-time. [2]

Figure 1. Traditional Meter

As shown in fig1.In apartments, the energy meter is far away from the residents. An LCD display is placed in each residential house in the apartment to inform about the messages regarding the power cut, energy consumption on daily basis, billing details and an alarm for the critical limit indication.[4]

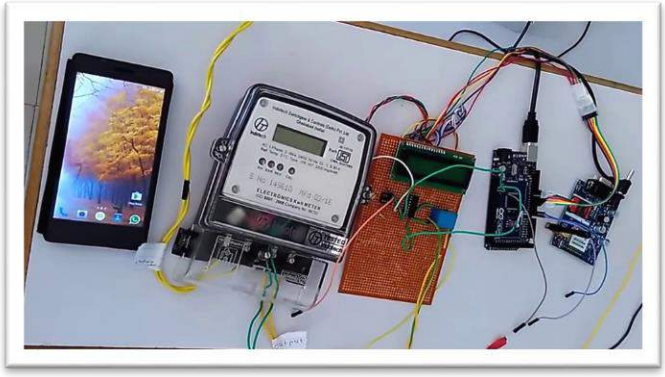


Figure 2. Proposed Iot based Smart Electric Meter

These features are implemented using the Arduino micro controller and a GSM module [1]. Esp8266 chip is the Wi-Fi module which helps in transferring the data through the internet. In case, if internet does not work, a GSM is used for backup purposes.

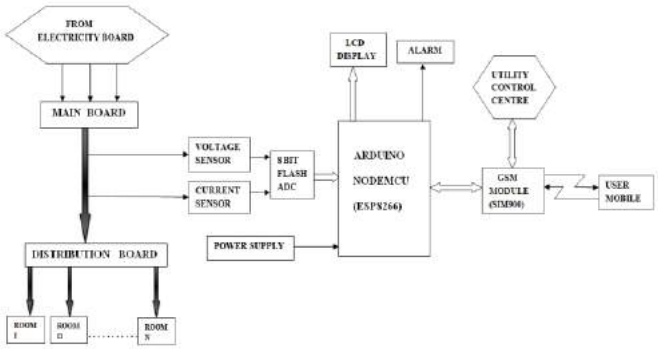


Figure.3. Functional block diagram of the Hardware module

2. RELATED WORK

Author S. Visalatchi ET. all in year 2017 published A new concept of electric meter will be discussed, where maximum demand of energy of a consumer will be indicated in the meter used by the consumer.[4]After exceeding the maximum demand, the meter and hence the connection will automatically be disconnected by an embedded system inserted in the meter itself.

Author P.V. Santhosh Roja, et.all. in year 2013 GSM MODULE SIM 300 is used to produce communication between load circuit and utility side[6]. We actually have used max232 along with DB9 connector to interface it.

Author Gobhinath.S,.et. all in year 2016 presented The system consists of the electricity meter which measures the electricity bill and informs the consumer about the number of units consumed and associated costs with it.[3] The microcontroller coordinates the whole system with the help of its different components connected to it. The design of Smart Energy meter using GSM technology can make the users to pay for the electricity before its consumption.

3. PROBLEM STATEMENT

In India, the current electricity billing system is completely manual. The electric meters are situated in the houses, offices and factories etc.

The energy meter reading is collected by meter readers on their fortnightly or monthly visits to the premises.

This system has disadvantage of appointing meter reader to take the meter reading, effects consumer privacy etc.

In this direction this project undertakes the meter reading without human intervention.

4. PROPOSED SYSTEM

4.1 Flow Chart of Proposed System

The existing model is a time-consuming process and it needs a lot of labour. The proposed system eliminates the need of labour and it is a cost efficient and a time saving process

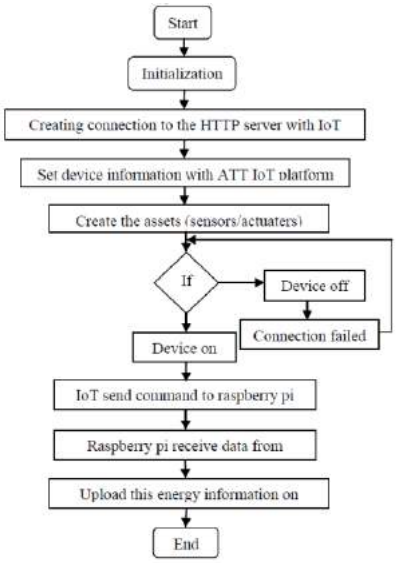


Figure 4. flow chart of proposed system

The proposed system gives the information about the energy consumption on daily basis, billing and payment through IoT, pre-intimation of shut down details, alert systems when the energy consumption exceeds beyond the critical limit and the disconnection of power through a message when the residential are out of station to prevent the wastage of energy.

4.2 SYSTEM DESCRIPTION

In this system, a unique Id is given for each energy meter. This unique Id number is interlinked with the customer's unique mobile Id number. It constantly monitors the energy meter. The energy consumption from each house is sent to the control station through the web server and the billing and power cut details are sent from the control station to the residential energy meter.

4.2.1 Energy consumption on daily basis

In the existing electric meter, the meter shows the energy consumed from the date of installation. In this system, the daily energy consumed is calculated using the Arduino microcontroller and it is displayed in the LCD. It is also communicated to the consumer's mobile using IoT.

4.2.2 Billing and Payment through IoT

The billing detail for the energy consumed is communicated to the consumer monthly through the webserver using IoT and as a message through the GSM module.

The payment is also made through the web server. This helps in eliminating the manual dependency to collect the reading.

4.2.3 Alert Systems

Our electricity billing system has a format that when the energy is consumed beyond the certain limit, then it indicates that there might be a electricity theft and alerts the consumer to check it.

4.2.4 Proposed Block Diagram

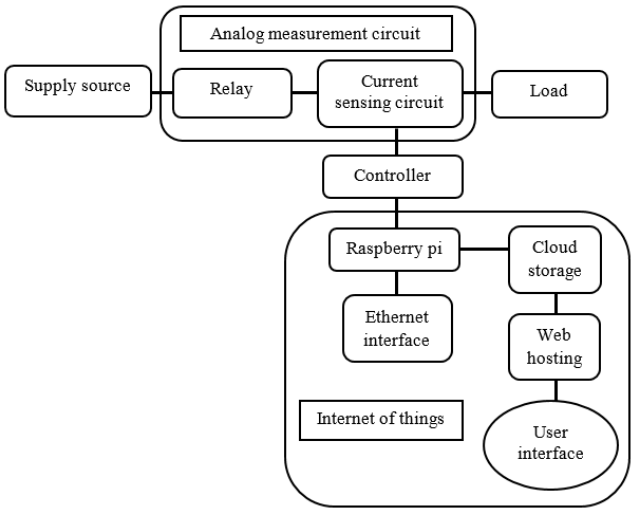


Fig. 5. Proposed block diagram

Arduino uno microcontroller is used which is based on ATmegha328. After receiving measured current data and input voltage from analogy measurement circuit we have to calculate the power. The programming is done in Arduino software (IDE). This is the open source platform. This board also acts as network traffic controller. It receives the measured data from sensor and send calculated data to Arduino . The instantaneous output is provided to the current flowing through the terminals by ACS712.

4.2.5 Internet of things

Arduino is the key learning platform for IOT. Arduino is connected to the controller and the internet. Once the Ethernet connection is set then login to raspberry pi to start developing IoT platform. Python programming language is used. Raspberry pi send energy information to the internet connected server for monitoring energy use information of individual device. Cloud storage and web hosting is used for storing real-time energy data and allows other programs to access and monitor the data. The control signal sends to raspberry pi using same server to turn on/off the device. These are the various steps to build IoT platform as shown in flow chart below.

5. RESUT ANALYSIS

The experimental setup for the system is as shown in fig. 3The ACS 712 current sensor gives precise current measurement for both AC and DC signals. These are good sensors for metering and measuring overall power consumption of systems. The ACS712 current sensor measures up to 20A of AC current.

- Project is focused on the government's plan to turn the major cities of the country into smart cities.
- The project can be further extended to detect the electricity meter tampering.
- A smart app can be designed to provide various alerts based on the readings from the device.
- A unified can be provided to the customers for both viewing the electricity usage and a platform to pay the bill online following the digital India initiative.
- In one case the service provider can evaluate the bills which are not paid and can disconnect the electricity connection remotely.

6. CONCLUSION

Arduino and GSM based Smart Electric Meter for advanced metering and billing system is built which is able to read and send data via wireless protocol using GSM technology through GSM modem, capable of managing the meter as well as the line connection. However, this project needs more modification for more reliable and higher degree of satisfaction and safety.

For GSM module the network coverage of the SIM used is one might be bit complicated due to network strength issues.

7. REFERENCES

[1]. Himshekhar Das and L.C. Saikia, "GSM enabled smart energy meter and automation of home appliances", Institute of Electrical and Electronics Engineers (IEEE), 2015.

[2]. V.Preethi and G. Harish, "Design and implementation of smart energy meter", Institute of Electrical and Electronics Engineers (IEEE), 2016. 188International Conference on Power, Energy, Control and Transmission Systems (ICPECTS). Authorized licensed use limited to: Middlesex University. Downloaded on September 04,2020 at 01:45:08 UTC from IEEE Xplore. Restrictions apply.

[3]. Md.Masudur Rahman, Noor-E-Jannat, Mohd. Ohidul Islam and Md. Serazus, "Arduino and GSM based smart energy meter for advanced metering and billing system", Institute of Electrical and Electronics Engineers (IEEE), 2015.

[4]. S. Visalatchi and K Kamal Sandeep, "Smart energy metering and power theft control using Arduino & GSM", Institute of Electrical and Electronics Engineers (IEEE), 2017.

[5]. Gobhinath.S, Gunasundari.N and Gowthami.P, "Internet of Things (IoT) Based Energy Meter", International Research Journal of Engineering and Technology (IRJET), 2016.

[6]. P.V. Santhosh Roja, B. Kiran Babu and V. Samson Deva Kumar, "Wireless Energy Meter and Billing via SMS", International Journal of Science and Research (IJSR), 2013.

[7]. Vanishree k Rao and Sri G N Madhu, "GSM based Energy Meter Reading and Billing", International Journal of Science and Research (IJSR), 2016.

[8]. Pritee Mahajan, Sneha Tatia and Prachi Jadhav, "Smart Meter: An IOT Based Meter for Live Electricity Monitoring and Bill Payment", International Journal of Electrical, Electronics and Computer Systems (IJEECS),2017

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Identify the Hacker Using IDS And Prevent the Hacker Using IPS to secure the Cloud Data

Geetanjali Pandey[1], Maithili Gavli[1], Shruti Khaire[1], Pragati Mote[1], Prof. Vandana Chavan[2]

[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

[2]Professor Computer Engineering, Dr. D. Y. Patil School of Engineering Lohegaon, Savitribai Phule Pune University, Pune, Maharashtra, India

ABSTRACT

Data-Security generally refers to the protective measures of securing data from unapproved access and data corruption throughout the data lifecycle. It measures not only helps avoid data breaches but also shields your organization against unnecessary financial costs, loss of public trust and potential threats to brand reputation and future profits too. Nowadays, the data is stored in the cloud. Thus, Cloud-Computing is the delivery of different services throughout the internet. These resources include tools and applications like data storage, servers, databases and networking. As long as an electronic device has access to the web, it has access to the data and software programs to run it. Cloud storage technology develops very fast, and cloud storage security technology is facing unprecedented

Challenges. However, cloud storage security is not just a technical issue [5]. Now, in the fifth generation increase in the use of cloud computing, lead to the demand of CLOUD-SECURITY.

Cloud-security have security principles applied to protect the data, applications and infrastructure associated within the cloud computing technology. Thus, we are developing an application to secure the cloud. The evaluation system includes security scanning engine, security recovery engine, security quantifiable evaluation model, visual display module and etc. The security evaluation model composes of a set of evaluation elements corresponding different fields, such as computing, storage, network, maintenance, application security and etc [4].

In order to effectively manage the networks for administrators within limited time and energy, we develop a hierarchical framework which detects the malicious attacks and prevent our data from that attacks. Thus, in our application we are using two algorithms, firstly IDS (Intrusion Detection System) to detect the attack, provide the information of the hacker to the administrator and the second algorithm used is named as IPS (Intrusion Prevention System) to prevent our data from the hacker. We are also going to retrieve the data of the hacker by using support vector machine (SVM).

Keyword—Data security, Cloud-Computing, Cloud security, IDS, IPS, SVM.

© 2020 IJSRST | Print ISSN: 2395-6011 | Online ISSN: 2395-602X | Published : 18/12/2020 - 85-90

I. INTRODUCTION

Capturing and analysing the abnormal behaviour is one of the most critical issues in keeping a network, data centre or cloud under control. Intrusion Detection System (IDS) and Intrusion Prevention System (IPS) are regarded as the most important techniques in security management [9]. We will develop a hierarchical framework to perform high threat mining and ranking based on their processing urgencies, in turn to reduce the operating difficulties for the network administrators. We have seen that personal computer's data and the cloud data are hacked due to less security provided by the user. This Data and the information is hacked or changed by the hacker, so we need to recover the hacked data or the retrieved data. In the existing system, there is no application to identify and detect the hacker. So in the current system, we use IDS and IPS techniques for detecting and preventing the data from the hacker.

IDS is a system that monitors network traffic for suspicious activity and issues alert when such activity is been discovered. It is a software application that scans the whole network or a system for harmful activity or policy breaching. A good intrusion detection system requirements for the highest possible detection rate and false alarm rate as low as possible due to intrusion detection in user behavior mainly as a data format, so the core problem is how to correctly and efficiently handle the data collected, and reach a conclusion.

An IPS is a system that has the ability to detect attacks, both known and unknown, and prevent the attack from being successful. Thus, Intrusion Prevention System is also known as Intrusion Detection and Prevention System

We are also retrieving the cloud data using SVM. SVM is a supervised machine learning algorithm and can be used for both classification and regression challenges. However, it is mostly used in classification problems.

Thus, in the proposed system, we are aiming to provide the security to our data stored in the cloud server, so that we can prevent our data from any malicious activity.

II. LITERATURE SURVEY

Table 1 – Literature Survey

Sr. No	Paper	Remarks
1.	Data Mining Based Intrusion Detection System application	Aimed on operating principle of IDS based on data mining [1].
2.	Design of a new Intrusion Prevention System of application	Introduction of IPS and discuss of the various threats To prevent them [2].
3.	Design of the Intrusion Detection System Based on Multi-Agents in the Ecommerce System	The system structure of IDS based on the Mobile Agent is proposed and then design of the MAIDS system [3].
4	One quantifiable security evaluation model for cloud computing platform	Quantifiable security evaluation system for different clouds that can be accessed by consistent API [4].

5.	Study on Data Security Policy Based On Cloud Storage	This paper is to achieve data security for cloud storage and to formulate the corresponding cloud storage security policy [5].

III. ALGORITHMS

1] IDS (Intrusion Detection System)

In this system we present Genetic Algorithm approach which can efficiently detect the various types of network intrusions. Genetic Algorithm is used to optimize the search of attack and by combining the IDS with genetic algorithm increase the performance of the detection rate of the Network Intrusion Detection Model and reduces the false positive rate. Intrusion Detection System (Intrusion Detection System) IDS surface it looks like network monitoring and alarm devices, a kind of observation and analysis of network attack has occurred, and to send a warning before the attack, and then do a corresponding counter-measures to reduce the huge losses the device may occur[1]. IDS is a system that monitors network traffic for suspicious activity and issues alert when such activity is been discovered. It is a software application which scans whole network or system for harmful activity or policy breaching. It monitors and analyses the user and system activities. It performs auditing of the system files and other configurations of the operating system. It assesses integrity of the system and data files. It conducts analysis of known attacks based on patterns. It detects errors in system configuration. It detects and alerts if the system is in danger. Intrusion detection system for detecting an attempt to undermine the integrity of computer resources, authenticity and availability of software behaviour, it can real-time monitoring system activities, real-time discovery of aggressive behaviour and take appropriate measures to avoid or minimize the occurrence of attacks generated by attack hazard[1].

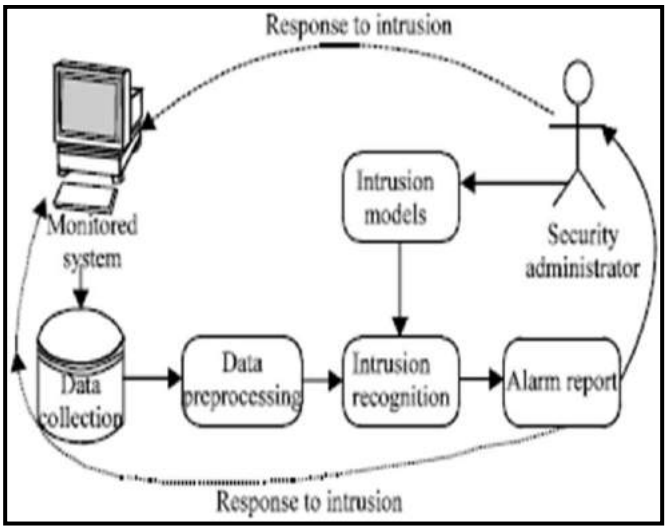


FIG 1 - IDS

2] IPS (Intrusion Prevention system)

An IPS is a system that has the ability to detect attacks, both known and unknown, and prevent the attack from being successful. Thus, Intrusion Prevention System is also known as Intrusion Detection and Prevention System. Intrusion Prevention System (IPS) is an important supplementary for security management [9]. Basically an IPS is a firewall which can detect an anomaly in the regular routine of network traffic and then stop the possibly malicious activity. The main functions of intrusion prevention systems are to identify malicious activity, log information about this activity, report it and attempt to block or stop it. Intrusion prevention is to detect and identify any abnormal user behaviour which aims at destructing information security and / or stability through the network [2]. Network-based IPSs create a series of choke points in the enterprise that detect suspected intrusion attempt activity. Placed inline at their needed locations, they invisibly

monitor network traffic for known attack signatures that they then block. These systems don't reside on the network per se but rather on servers and individual machines. They quietly monitor activities and requests from applications, weeding out actions deemed prohibited in nature. These systems are often very good at identifying post-decryption entry attempts. These IPSs scan network packets, looking for signatures of content that is unknown or unrecognized or that has been explicitly labelled threatening in nature. Intrusion Detection System (Intrusion Detection System)IDS surface it looks like network monitoring and alarm devices, a kind of observation and analysis of network attack has occurred, and to send a warning before the attack, and then do a corresponding counter-measures to reduce the huge losses the device may occur[1].

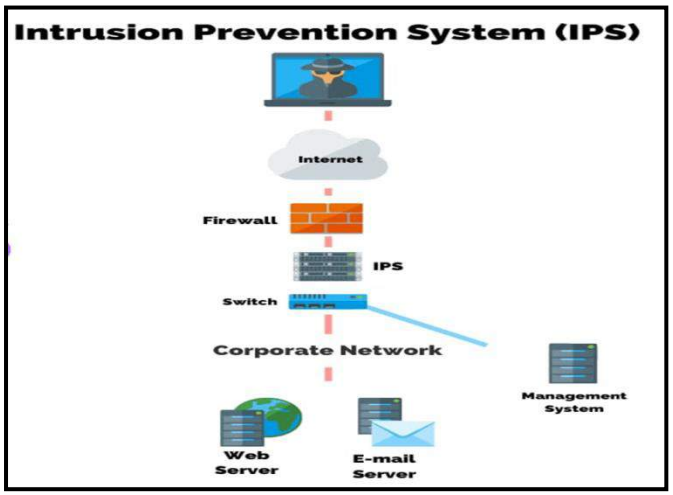


FIG 2- IPS

3] SVM (Support Vector Machine)

SVM is a supervised machine learning algorithm which can be used for both classification and regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well. SVM's are very good when we have no idea on the data. Works well with even unstructured and semi structured data like text, Images and trees. The kernel trick is real strength of SVM. With an appropriate kernel function, we can solve any complex problem. It scales relatively well to high dimensional data.

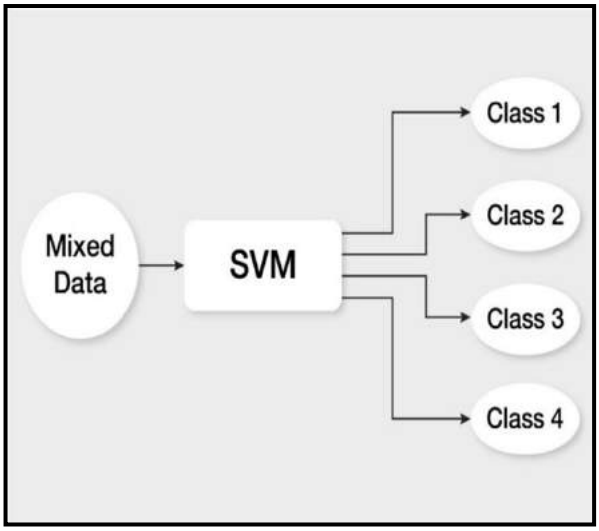


FIG 3- SVM

IV. EXISTING SYSTEM

In existing system there is no computerizes system to identify intrusion detection attack in your personal computer or laptop. A hacker can easily change your personal database or hack our personal database. But we cannot identify them so we can't understand who is stolen our data. So in proposed system we are trying to give security to our data and stored out data in cloud server so hacker cannot identify the data storage location. The existing network intrusion detection research is mostly concentrated on the wired network, the intrusion detection research on wireless sensor networks is relatively little [2].

© 2020 IJSRST | Print ISSN: 2395-6011 | Online ISSN: 2395-602X | Published : 18/12/2020 - 85-90

V. MATHEMATICAL MODEL

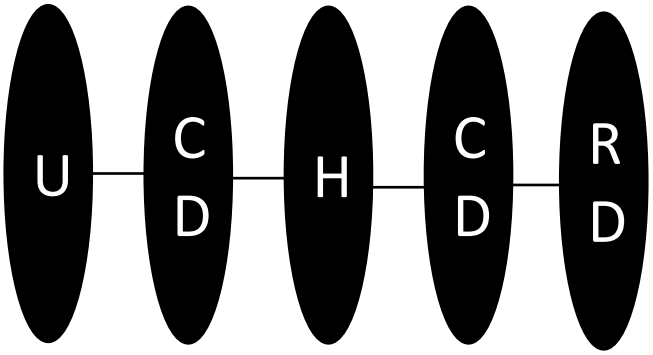


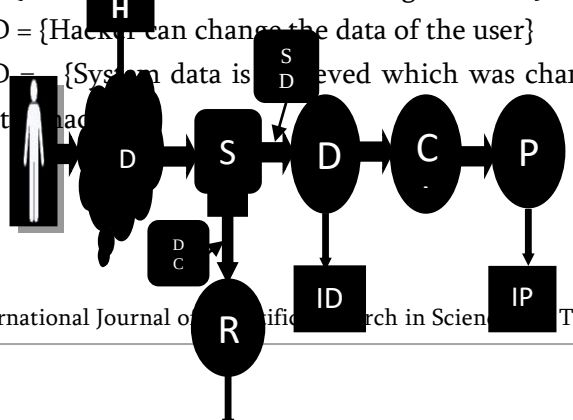
FIG 4 - MODEL

Where,
U =User stores data on Cloud
CD=Data stored on cloud server
H =Hacker can make login attempt
CD =Hacker changes the data
RD =Retrieve the original data.
Above mathematical model is NP-Hard
Because sometime result is not accurate.

Input: Hacker can make login attempt on the user's Pc.
Output: System then captures the hacker's face, retrieve the data and system is blocked.

Let us consider, H as hacker who can make login attempt on user's PC and change the data.
H = {U, CD, CD}
Where,
U = {User can upload data on cloud server.}
CD = {Cloud server store the user's data}
CD = {Hacker can change the data of user}

U = {H, CD, RD}
Where,
H= {User [illegible]ves hacker's face image via mail}
CD = {Hacker can change the data of the user}
RD = {System data is [illegible]eved which was changed by t[illegible]ac



Functions: Functions implemented to get the businessman original data and detect the hacker face. Functional relations: 1] Hacking, 2] Security, 3] IPS, 4] IDS.

Success Condition: Successfully algorithm implementation and proper input
Failure Condition: 1. huge data can lead to more time consumption to get the information.2. Hardware failure.3. Software failure.

Space Complexity: The space complexity depends on Presentation and visualization of discovered patterns. More the storage of data more is the space complexity.
Time Complexity: Check No. of patterns available in the database = n. If (n > 1) then retrieving of information can be time consuming. So the time complexity of this algorithm is O (n ^n).

VI. SYSTEM ARCHITECTURE

If hacker is trying to hack the data of our system , we will catch the face of the hacker if login attempt fail at the first time.In the second time if attacker changed the data on our PC, then our system will retrieve the previous data using support vector machine. If the hacker attempted for the third time to hack the system then we will block the system, and we will not provide any login option for the hacker.

FIG 5 – SYSTEM ARCHITECTURE

VII. Advantages of System Architecture

[1] Using IPS and IDS algorithm system can provide the security to the user's important information and data.

[2] SVM algorithm can recover the user's important information and data which is changed or modified by the hacker.

[3] This is reliable system.

[4] This system can prevent the hackers from hacking.

[5] When hacker trying to hack the user's important information and data then system send the email of the hacker's image to user, because of this email system immediately alert the user.

[6] Replace a Human Monitoring Your Network 24x7.

VIII. APPLICATIONS

[1]. Small business: The reason being, many large companies have the infrastructure in place to guard against cyber-attacks. Small businesses, however, either don't have the proper resources to thwart an attack or they don't take cybersecurity as seriously as they should.

[2]. Healthcare: The healthcare industry is another prime target for ransomware attacks because of the sheer amount of patient data stored by healthcare entities. Health information is some of the most valuable data on the dark web because it can be used to commit insurance fraud.

[3]. Higher Education : When you think of potential targets for hackers, colleges and universities probably aren't the first to come to mind, however, the higher education industry is another mecca of personal data. From social security numbers, addresses and passwords to loan and bank information, it's no wonder attacks on colleges and universities are becoming more prevalent.

[4]. Energy: Last, but by no means least, is the energy sector. Here, things like the electric power grid and power generation facilities are controlled by technology and communication systems that could be disrupted, hacked or taken over during a cyber-attack to put our economy in serious danger.

IX. CONCLUSION

In order to effectively manage the networks for administrators within limited time and energy, we develop a hierarchical framework to secure the data of the user by detecting and preventing any malicious attack. With the help of IDS and IPS our data is highly secure. We can also get image of the person who is unauthorized accessing our data. If our data is hacked then we can also retrieve it. We can also block the system if hacker is repeatedly trying to attempt the login. Thus, we find that the accuracy of our proposed method is larger than 97%, the analysis results verify that our proposed methods is more effective compare with other methods[9].

X. FUTURE WORK

[1] Malware targeting virtual machines: "Many breeds of malware today can detect if they are running within virtual machines and make

adjustments or shut down altogether in order to evade detection, but only a few proof of concept viruses have actually attempted to break free into the host machine," explained Fred Couchette, senior security analyst at Approvers. "We expect to see more of these in the near future."

[2] ATM-like hardware hacks: "We've seen criminals physically walk in to stores and replace credit card terminals with working replacements that had been modified to contain a 3G modem, which transmitted payment details directly back to them," said Lyne. "This high scale, intelligent hardware hacking demonstrates that the threat is not just impacting the conventional PC."

[3] RAM scraping: "For years everyone has been locking down databases since they are the source of information, but now hackers that can breach a server can get an application less than 1MB in size on the server and capture all the data as it is written to RAM before it goes to a database," said Chris Drake, CEO of Fire Host. "An application like this can also capture data (such as credit card numbers) that don't even go into a database, but that are processed by a third party provider. RAM scraping will be a huge concern as it gains more popularity among the hacker crowd.

XI. REFERENCES

[1]. FAN Yaquina College of Communication. Engineering, Jilin University Changchun 130012, china fanyaqin_joy@163.com, Paper Name- Data Mining Based Intrusion Detection System in VPN Application.

[2]. Huangzhong Hua School of Mechatronical Engineering Beijing Institute of Technology Beijing China fengqingjuan@foxmail.com, Paper Name- Design of a new Intrusion Detection System of WSNs.

[3]. Yabo Li College of Business Administration of HUNAN University lyb412@sina.com, Paper Name- Design of the Intrusion Detection System Based on Multi-Agents in the Ecommerce System.

[4]. Aobing Sun1,2* , Guohong Gao1 , Tongkai Ji1,2 , Xuping Tu1,2, Paper Name-One quantifiable security evaluation model for cloud computing platform.

[5]. DIAO Zhe, WANG Qinghong, SU Naizheng, ZHANG Yuhan University of International Relations, Beijing, China, Paper Name-Study on Data Security Policy Based On Cloud Storage

[6]. Salah S, Maciá-Fernández G, Díaz-Verdejo J E, "A model-based survey of alarmcorrelation techniques," Computer Networks, vol. 57, pp. 1289- 1317, 2013.

[7]. Dwivedi, Neelam, and A. Tripathi, "Event Correlation for Intrusion Detection Systems," IEEE International Conference on Computational Intelligence & Communication Technology IEEE, pp. 133-139, 2015.

[8]. Zhang S, Gao Y, Zhang M, et al, "The Study of Network Security Event Correlation Analysis Based on Similar Degree of the Attributes," Digital Manufacturing and Automation (ICDMA), pp. 1565-1569, 2013.

[9]. Yongwei Meng. Tao Qin, Member, IEEE. Yukun Liu, Student Member, IEEE and Chao He.Paper Name-An Effective High Threating Alarm Mining Method for Cloud Security Management.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Automated Website Development

Tushar Gangurde[1], Shivaraya Patil[1], Vaibhav Patil[1], Akshya Mahale[1]
[1]Computer Science, Savitribai Phule Pune University / Dr DY Patil School of Engineering, Pune, Maharashtra, India.

ABSTRACT

A website helps a business to grow by using different marketing strategies. This paper describes a novel approach to develop a website by just providing the text (description of the website) or an image as input. Using Text Input it will suggest template (screenshots) after identifying the theme of the site inferred from the input. Those templates are converted into code for further customizations for their personal use. Current problem was that a web developer will take more than 15 days only to just make the basic structure of a website. This issue is resolved by our work which will generate the complete code of the webpage/ website in less amount of time. In this paper, it will tokenize each word to find their synonyms and then mapped it with root words for the theme identification and uses deep learning model to convert templates into code.

Keywords: Root words, Theme, Synonyms, code

I. INTRODUCTION

To survive in the digital world, website becomes the basic requirement of each business to represent itself in digitalized world, big or small. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored.

If a business does not have a website, they are losing the number of great opportunities for their business. A website helps a business to grow by using different marketing strategies. So, deploying a high end and user interactive websites are important and should be done in minimal time as possible.

What makes a website?

- A website contains HTML code which acts as its Skeleton for website
- Then comes CSS (Cascading Style Sheet) and JS (JavaScript) which is as the name says used for design the website

The user input can be in any structure of the sentence from which tokens will be extracted and their synonymies will be mapped to identify the theme from root word. After identifying the theme, Suggestions will be shown and that suggestion will be feed into Convolution neural network (CNN) after that features will be extracted and feed into Long Short Time Memory (LSTM) which will generate DSL Tokens. Those tokens will be compiled to generate code.

Paragraph Segmentation:
first step is to break the word in segments of words which is called tokens in NLP. Using Word tokenize package from NLTK. The tokens contain English words, every token is extracted from the sentence of words.
Word Analysis is now after the sentence is converted into tokens. Then part of speech tagging is done for each word. Words containingonly Noun and adjective are extracted from the tokenized.

Only Noun and adjective are used to detect theme because:

 Noun contains words like objects, actions qualities and state of existence

 The adjective is useful in identifying parts of speech with the noun which is mainly the motive of word analysis.Word Analysis Now after the sentence is converted into tokens. Then part of speech tagging is done for each word. Words containing only Noun and adjective are extracted from the tokenized.
Only Noun and adjective are used to detect theme because:

 Noun contains words like objects, actions qualities and state of existence

 The adjective is useful in identifying parts of speech with the noun which is mainly the motive of word analysis.

Word Mapping and Template suggestion

Next Step is to map the words with their original word which is already present in the list to map each word we need to find synonyms for each word by using WORDNET which is a lexical database for English language, by using Wordnet.synset() function to find the synonyms of the word.
After finding synonyms the words will be mapped with root words for theme after mapping. Templates are already classified which template comes in which theme on
the base of the identified theme. Templates will be suggested to the user.

CNN

In deep learning, to analyze visual images the mostw common class of neural networks used is convolutional neural network (CNN, or ConvNet). CNN require minimal preprocessing because it is designed by using a variation of multilayer perceptron.[1] Due to shared-weights architecture and characteristics of translation invariance they are known as shift invariant or space invariant artificial neural networks (SIANN). Inspiration from biological processes [4], CNN develops a network of neurons by proving the connectivity pattern in the way that it looks like the pattern used in animal visual cortex. In receptive field, each individual cortical neuron will respond only in restricted region to the stimuli of the visual field. Different neurons show partially overlapping in the receptive fields in such a way that they will cover the entirevisual field.

LSTM

Recurrent neural networks (RNN) have the units as long short-term memory (LSTM). LSTM network consist of various LSTM units to build a recurrent neural network. Various components of LSTM unit are cell, a forget date, an input and output gate. All the 3 gates control the transmission of information in and out of the cell and then the cell stores the values on arbitrary intervals of time. To classify the data, process it and perform predictions which are on the basis of time series data, most commonly networks

used are LSTM networks because in important time series events there can be uneven or unknown duration lags. While training the traditional recurrent neural networks various vanishing and exploding gradient problems occur and LSTMs are developed to handle such problems. In various applications, relative insensitivity to gap length makes LSTM better than recurrent neural networks, hidden Markov models and other sequence learning methods.

II. Our Approach

The proposed system is done in various steps like text segmentation, tokenizing, part of speech tagging, word map and theme, suggesting, CNN, LSTM and decoder as shown in Figure 1.

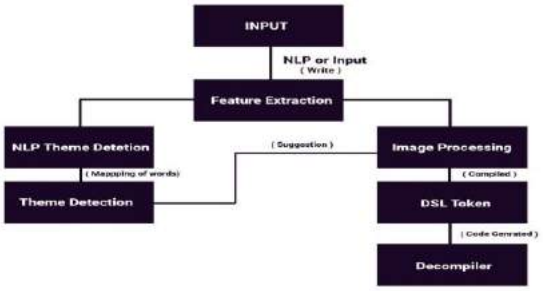


Figure 1:.Steps of approach

2.1 Input Image and Text

In this system the user can input text in two ways:

- User can upload images of UI to convert UI to code.
- User can type text description which will be used to suggestion and from that user can select the type of UI they went from the suggestion and that will generate code.

Figure 2: User Input

2.2 Feature Extraction

Both the methods of input have feature extraction Algorithms which extracts characteristics from both inputs each of them has their own feature extraction algorithm.

2.2.1 Image input uses CNN as a feature extraction algorithm to get characteristics of input. CNN is widely used in computer vision problems because of its topology which allows them to extract minor details from the input. We used convolutional neural network which behaves like an encoder and performs unsupervised learning by comparing an input image to a fixed-length vector which was already made learn to system.

2.2.2 Text input uses NLTK python library to firstly tokenize the inputs from which Noun and Adjectives are extracted from the tokenized data.

EXAMPLE:

INPUT: “Need a website with red color navigation panel and black background”

OUTPUT:
[„'Need'„website„with„„red„„colour„„navigation„„panel
„„
black„ background)

2.3 Theme Detection

After Tokenizing these phases come into play which Entity Recognition or Chunking which extracts Noun and Adjectives from tokenized data which are called Chunked data from that data synonyms are extracted and mapped with root words.

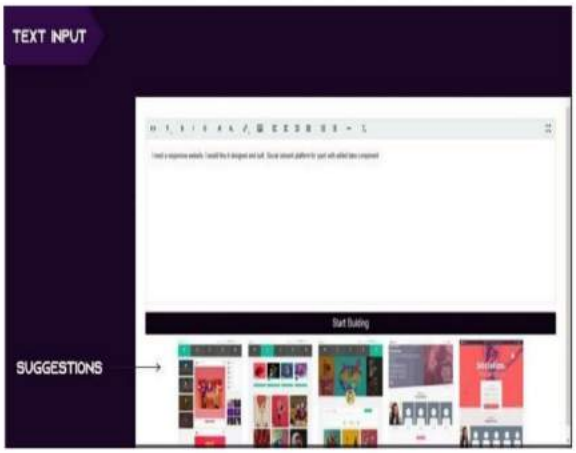


Figure 3: Theme Detection from text Input

2.4 Image Processing and DSL Token generation:

After the extraction of features from the image using CNN, we used DSL tokens to describe UI Components. To find different graphical components and their relation between each other DSL Token generation is used. DSL reduces the size of search space by reducing the total number of tokens of vocabulary of the DSL. By providing a discrete input, our system language model performs the modeling at token-level which uses one-hot encoded vectors, eliminating the need for word embedding techniques such as word2vec.

2.5 Decoder

By using supervised learning method, model is trained by inputting an image I and xt is a contextual sequence of X of T tokens, $t \in \{0 \ldots T-1\}$ as inputs; and xT token is taken as the target label. Input image I is encoded into vector representation p by using CNN-based vision model. LSTM Model is used to encode the input token xt into an intermediate representation qt which allows the model to concentrate more on certain type of tokens and less focus on others.

Each LSTM layer consists of 128 cells and the first language model consist of two such layers. A single feature vector rt is formed by the concatenation of p(vision encoded vector)and qt(language encoded vector) which is then given as input to a second LSTM-based model. This model will decode the image by mapping with various models like vision and language model. This makes the decoder able to learn to map the relation between the
objects identified in GUI image provided as input and the
tokens present in DSL code. In out decoder each LSTM layer consists of 512 cells and it is implemented as the combination of two LSTM layers.

The architecture discussed above can be represented in
mathematical form as:

$p = CNN(I)$
$qt = LSTM(xt)$
$rt = (q, pt)$
$yt = softmax(LSTM0\ (rt))$
$xt+1 = yt$

© 2020 IJSRST | Print ISSN: 2395-6011 | Online ISSN: 2395-602X | Published : 18/12/2020 - 91-97

III. Result

This work will generate code for UI input provided by the user. Table 1 shows input provided by user as an image and code is the output generated by the system. Table 2 shows input provided by user as text and themes are generated by the system as output.

Figure 4:Input 1

Figure 6: Input-2

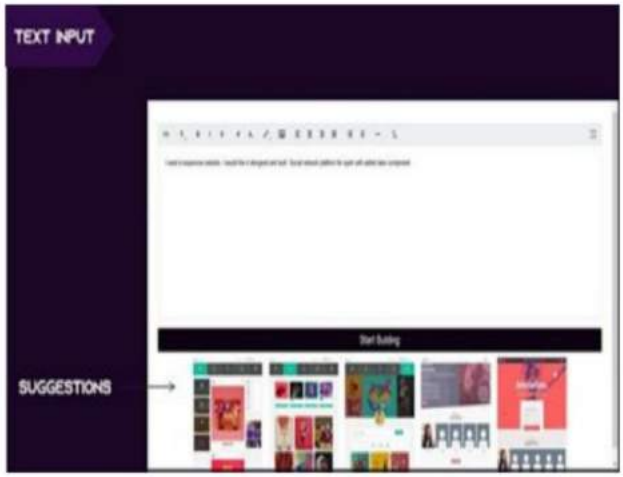
```
START <!DOCTYPE html>
<html lang="en" dir="ltr">
<head>
<title>Basic 88</title>
<meta charset="iso-8859-1">
<link rel="stylesheet" href="styles/layout.css" type="text/css">
<!--[if lt IE 9]><script src="scripts/html5shiv.js"></script><![endif]-->
</head>
<body>
<div class="wrapper row1">
  <header id="header" class="clear">
    <div id="hgroup">
      <h1><a href="#">Basic 88</a></h1>
      <h2>Free HTML5 Website Template</h2>
    </div>
    <nav>
      <ul>
        <li><a href="#">Text Link</a></li>
        <li><a href="#">Text Link</a></li>
        <li><a href="#">Text Link</a></li>
        <li><a href="#">Text Link</a></li>
        <li class="last"><a href="#">Text Link</a></li>
      </ul>
    </nav>
  </header>
</div>
</body>
</html>
```

Figure 5: Output-1

```
<body>
<div class="wrapper row2">
<!-- main content -->
<div id="homepage">
    <!-- Services -->
    <section id="services" class="clear">
      <article class="one_third">
            <figure><img src="images/demo/290x180.gif"
width="290" height="180" alt="">
        <figcaption>
          <h2>Indonectetus facilis</h2>
            <p>Nullamlacus dui ipsum conseque loborttis
non euisque morbi penas dapibulum orna.</p>
            <footer class="more"><a href="#">Read More
&raquo;</a></footer>
        </figcaption>
      </figure>
    </article>
    <article class="one_third">
            <figure><img src="images/demo/290x180.gif"
width="290" height="180" alt="">
        <figcaption>
          <h2>Indonectetus facilis</h2>
            <p>Nullamlacus dui ipsum conseque loborttis
non euisque morbi penas dapibulum orna.</p>
            <footer class="more"><a href="#">Read More
&raquo;</a></footer>
        </figcaption>
      </figure>
    </article>
</section>
</body>
</html>
```

Figure 7: Output-2

TABLE 2: Text Input and OUTPUT.

INPUT
I need a responsive website. I would like it designed and built. Social network platform for sport with added dare component
OUTPUT


IV. CONCLUSION

This proposed system have discussed the theme detection technique and suggesting themes to user and generating code from the selected template. Current problem was that a web developer will take more than 15 days only to just make the basic structure of a website. This issue is resolved by our work which will generate the complete code of the webpage/ website in less amount of time. This system can be used by anyone to make a website from just a text input or screenshot of a website to generate code from it. It can be used to generate code for Android or IOS Applications from there GUI Screenshots to fully functional code. Our model is trained on very small dataset increasing the size dataset can lead to more accurate results. Our work is concerned with the generation of static website code. This work can be further extended to developing or generating the code for both static as well as dynamic websites.

V. REFERENCES

[1]. J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell. 2015. Long-term recurrent convolutional networks for visual recognition and description. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2625–2634.

[2]. F. A. Gers, J. Schmidhuber, and F. Cummins. 2000. Learning to forget: Continual prediction with lstm. Neural Computation, 12(10), 451–247.

[3]. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. 2014. Generative adversarial nets. In Advances in neural information processing systems, pages 2672–2680.

[4]. A. Karpathy and L. Fei-Fei. 2015. Deep visual-semantic alignments for generating image descriptions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3128–3137.

[5]. T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. 2013. Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems, pages 3111–3119.

[6]. W. Zaremba, I. Sutskever, and O. Vinyals. 2014. Recurrent neural network regularization. arXiv preprint arXiv:1409.2329.

[7]. Gurpreet Kaur, Prateek Agrawal. 2016. "Optimisation of Image Fusion using Feature Matching Based on SIFT and RANSAC", Indian Journal of Science and Technology, 9(47), pp 1-7.

[8]. H. Zhang, T. Xu, H. Li, S. Zhang, X. Huang, X. Wang, and D. Metaxas. Stackgan. 2016. Text to photo-realistic image synthesis with stacked generative adversarial networks. arXiv preprint arXiv:1612.03242.

[9]. A. Krizhevsky, I. Sutskever, and G. E. Hinton. 2012. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105.

[10]. D. Bahdanau, K. Cho, and Y. Bengio. 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.

[11]. A. L. Gaunt, M. Brockschmidt, R. Singh, N. Kushman, P. Kohli, J. Taylor, and D. Tarlow. 2016. Terpret: A probabilistic programming language for program induction. arXiv preprint arXiv:1608.04428.

[12]. W. Ling, E. Grefenstette, K. M. Hermann, T. Kocisk ˇ y, A. Senior, F. Wang, and P.Blunsom. 2016. Latent predictor networks for code generation. arXiv preprint arXiv:1603.06744.

[13]. L. Yu, W. Zhang, J. Wang, and Y. Yu. 2016. Seqgan: sequence generative adversarial nets with policy gradient. arXiv preprint arXiv:1609.05473.

[14]. S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee. 2016. Generative adversarial text to image synthesis. In Proceedings of The 33rd International Conference on Machine Learning, volume 3.

[15]. K. Xu, J. Ba, R. Kiros, K. Cho, A. C. Courville, R. Salakhutdinov, R. S. Zemel, and Y. Bengio. 2015. Show, attend and tell: Neural image caption generation with visual attention. In ICML, volume 14, pages 77–81. View publication stats.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Voice-Based Intelligent Virtual Assistant for Windows using Speech Recognition and Speaker Identification Technology

Vrushali Kolte[1], Samidha Jadhav[1], Kalyani Kasar[1], Prof. Ashwini Pandagale[2]

[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

[2]Professor, Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

Voice assistants developed by big companies like Cortana by Windows, Alexa by Amazon, Siri by Apple, and Google assistant by Google which performs the task ordered by the user. It can play audio-video, search for something, book flight tickets, can have fun conversations, etc. But the downfall of such technologies is the security issue as it stores its data on the cloud which can be retrieved by any technique and can be misused. Another issue is battery backup as most of the assistants do not have an inbuilt battery to work even in light cut-off situations. To overcome the above issues this project introduces Intelligent Virtual Assistant (IVA).

IVA – Intelligent Virtual Assistant which not only follows her boss's order but also gives the next response using her artificial brain. This response may advice, motivation, choices, legal actions, etc. It comprises of gTTS, pyttsx, AIML [Artificial Intelligence Markup Language], and Python-based state-of-the-art technology. The Internet has made information more accessible over a wide network, thus making it quicker and vaster, revolutionizing how people communicate in the world. The information available on the Internet about a given topic may be extensive, which helps in finding the solutions to day-to-day problems. And hence this project is based on forming more communication interactive models with the use of gTTs and AIML, facilitating the establishment of considerably smooth dialogues between the assistant and the users. Also, IVA stores its data in the user's PC this eliminates security problems.

IVA is also capable of recognizing the user's voice its gesture. This project not only gives logical or technical output but also an emotional one. IVA is a software agent that can assist people in many of their daily activities. It is capable of retrieving information from databases to give suggestions to people on performing different tasks, deploying a learning mechanism to acquire new information on user performance. It can make assistance more reliable and efficient by collecting information autonomously from objects that are available in the surrounding environment. This project also consists of voice-based user verification using a deep learning framework that recognizes the user's voice by its timbre and pitch. To make this idea feasible, IVA uses many searching protocols, artificial intelligence, machine learning, deep learning, etc.

Keywords: Cortana, Alexa, Siri, Google Assistant, gTTs, pyttsx, AIML, artificial intelligence, machine learning, deep learning.

© 2020 IJSRST | Print ISSN: 2395-6011 | Online ISSN: 2395-602X | Published : 18/12/2020 - 98-103

I. INTRODUCTION

The term virtual assistant was coined way back in 1950s even before Siri which was developed by Apple as a virtual assistant for android. The term virtual assistant or virtual personal assistant is an application program that can understand natural human language, speak natural language and complete electronic task for the end user [11].The main aim is to design a voice-based intelligent virtual assistant (IVA) that acts as a digital organizer to provide variety of services to its master with the use of various machine learning algorithms, which accept voice input, process it and provide desired output to user[12].Some virtual assistants are able to interpret human speech and respond via synthesized voices, Users can ask their assistants questions, control home automation devices and media playback via voice, and manage other basic tasks such as email, to-do lists, and calendars with verbal commands, provide Profile management, Reminders [13], etc.

This intelligence system is classified into three generations: First Generation based on Pattern Matching; Second Generation including techniques of Artificial Intelligence such as deep neural network; Third Generation indulging higher ordered, sophisticated pattern matching techniques which is mostly based on AIML, a markup language for chat-bots constructions which is based on XML[3].

Following are the technologies used in Virtual assistant:

A. Speech Recognition

Speech Recognition is the ability of machine/program to identify words spoken aloud and convert them into readable text.

The main building blocks of speech recognition system are: Signal preprocessing Feature extraction, Language model, Decoder and Speech Recognition[5].

B. Speaker Identification

Speaker identification is the computing task of validating a user's claimed identity using characteristics extracted from their voices.

The two main modules of speaker identification are feature extraction and feature matching. Feature extraction is used to differentiate speaker according to their pitch and tone. In testing phase first features are extracted and then they are matched with the speaker templates using feature matching module[9].

II. LITERATURE SURVEY

Here we have discussed the literature survey of some of the existing techniques.

Giancarlo Iannizzotto et al. [1] proposed A vision and speech enabled, customizable, virtual assistant for smart environments. In this paper, an architecture is created that can assist the user not just only by voice but also by taking vision of the users. This paper introduces an architecture for building vision-enabled smart assistants, provided with expressive and animated graphical characters and speech recognition and synthesis. The limitation of this system was lack of accuracy in recognizing speaker and

system can be trained for fall detection and anomalous behaviour.

Veton Këpuska and Gamal Bohouta [2] proposed Next-Generation of Virtual Personal Assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home). In this Proposal, an approach is proposed that will be used to design Next generation of Virtual personal assistants, increasing interaction between users and the computers by using the Multi-model dialogue system with techniques including gesture recognition, image/video recognition, speech and conversational knowledge base, General knowledge based. The advancement of the system is to achieve more accurate result and customer service home automation.

Neha Madhavi et al. [3] proposed JARVIS: An interpretation of AIML with integration of gTTS and Python. The system comprises of three modules first is based on Pattern Matching, second includes techniques of Artificial Intelligence and third includes indulging higher ordered, sophisticated pattern matching techniques, based on eXtensible Markup Language (XML). All this module has worked on achieving high accuracy of the system. But there are some limitations of this system that is AI should be transformed from stable AI to dynamic self-learning AI system which will be more potent to the system itself.

Artur Dovbysh and Vladyslav Alieksieiev [4] proposed Development and integration of speech recognition tools into software applications and an approach to improve of speech recognition quality. The purpose of the current research paper is to improve existing free to use tools of speech recognition and develop the approach to achieve both speech recognition engine and modules than bring recognition improvements with custom application. This architecture mainly includes speaker's diarization, Word correction text filter which includes Levenshtein's distance technique. To improve speech recognition efficiency, a primitive Python Neural network was implemented and using the TensorFlow training Library. The weak point of the system was differentiation of speakers' accents.

Ashok Kumar and Vikas Mittal [8] proposed Speech Recognition: A Complete Perspective. This paper introduces building blocks for general speech recognition system i.e. Signal pre-processing Feature extraction, Language model, Decoder and Speech Recognition. Challenges in speech recognition like environment factors, Vocabulary solves where overcome using above building blocks. Fields like local language or foreign language recognition with efficiency has to work still.

Sanket Shah et al. [6] proposed Speech recognition using neural network. This paper focuses on deep neural network which is a modern approach for recognizing the speech. Various neural network such as DNN, RNN and LSTM is discussed in the paper. The LSTM algorithm complexity per weight and time step is O(1). This is the extreme advantageous for high speed working of system and outruns the other approaches such as RTRL. The advancement

in this field is that it can be used in various languages depending upon the region being used.

Nisha [7] proposed Voice Recognition Technique: A Review paper. In this paper biometric technology is used to recognize a particular individual's voice. The speech waves of particular voice form the basis of identification of speaker. Voice has its unique characteristics called feature & the process of extracting these features from the individual voice is called feature extraction. Techniques used in this system are Linear Predictive Coding, Perceptual Linear Prediction, Relative Spectral Filtering and Mel Frequency Cepstral Coefficient. Features are used to train a classifier so that it can classify the words which are spoken by the subject. Various classifiers which are used in voice recognition system are Hidden Markov Model, Neural Network Model, Dynamic Time Warping and Vector Quantization. This paper provides review of various voice and speaker recognition systems.

David Snyder et al. [8] proposed Speaker Recognition for Multi-Speaker Conversations using X-Vectors. This paper introduces x-vectors, an embedded neural network that is very effective for both speaker recognition and diarization. Speaker diarization is the process of grouping segments of speech according to the speaker, and is sometimes referred to as the "who spoke when" task. This system solves problem of speaker recognition on multi-speaker conversation

Satyam P. Todkar et al [9] proposed Speaker Recognition Techniques: A review paper. This paper has aimed to identify the speaker from different speakers available or verify a particular speaker. The voice of every individual sound is different as they are attributed to different features that create the voice, this may be- pitch, length of the vocal tract, sound frequency etc. Algorithms used are Vector Quantization (VQ), Gaussian Mixture Model (GMM), Neural Networks (NN), Mel Frequency Cepstral Coefficients (MFCC), Linear Predictive Coding (LPC), GMM & Pitch Detection Algorithm, Hidden Markov Model (HMM). The system was able to cut off background noise using MFCC algorithm.

Garima Sood et al [10] proposed A Robust Speaker Identification System for Natural and Whispered Speech. This paper solves the main challenge lies in improving the robustness of the system in highly noisy environment. using, different identification algorithms for both normal and whispered speech have been compared to check the robustness. The extracted features have been classified with various classifiers such as Support Vector Machine, Fine K Nearest Neighbour (KNN) and Weighted KNN. Again, the limitation of the system was acquiring high accuracy.

Below are some of the highlighted researches in speaker recognition and speaker identification field:

Sr. No.	Author	Techniques used	Challenges	Constraints
1	Giancarlo Iannizzotto et al.	Mary TTS module Flite2 module	To make interaction more natural and attractive.	Fall detection or anomalous behaviour.
2	Gamal Bohouta et al	Next-generation of VPA model	To establish a lengthy conversation with user.	System can be more accurate and can be used in customer service home automation.
3	Neha Madhavi et al.	AIML and Google API	Interpretation of AIML.	To transform stable AI assistant into a dynamic self-learning AI system.
4	Artur Dovbysh et al.	Speaker Diarization	To improve speech recognition quality.	Differentiation in speaker's accent.
5	Ashok Kumar et al.	Linear predictive coding (LPC), PLP and Mel Cepstral Coefficients (MFCC)	Challenges in speech recognition like environment factors, Vocabulary solves.	Fields like local language or foreign language recognition with efficiency has to work still.
6	Sanket Shah et al.	Speech classification techniques.	To obtain highest accurate results.	Speech recognition can be used for various languages.
7	Nisha	ASR System	To gain highest result from various feature extraction techniques.	Access control or Transaction Authentication.

8	David Snyder et al.	Speaker diarization, x-vector, AHC.	Speaker recognition for multi-speaker conversation.	AHC can be replaced with alternative method.
9	Satyam P. Todkar et al.	Feature matching and extracting techniques.	To identify speaker from different speakers.	Useful in noisy environment.
10	Garima Sood et al.	SVM AND K-NN	Speaker identification for natural and whispered speech.	Real time training and useful with IOT devices.

III.CONCLUSION

This paper introduces voice-based intelligent virtual assistant specifically designed for windows operating system. In this system we have integrated both speech recognition and speaker identification technology. This IVA system uses speech, audio, video and other mode of communication to interact with human. Further this system can be used in different fields such as home automation, medical assistance, car automation, robotics and security access.

IV. REFERENCES

[1]. Giancarlo Iannizzotto, Lucia Lo Bello, Andrea Nucita, Giorgio Mario Grasso "A vision and speech enabled, customizable, virtual assistant for smart environments" Dept. for Cognitive Sciences, Psychology, Education and Cultural Studies (COSPECS).

[2]. Veton Këpuska, Gamal Bohouta Electrical & Computer Engineering " Next-Generation of Virtual Personal Assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home) " Department Florida Institute of Technology Melbourne, FL, USA

[3]. Ravivanshikumar Sangpal, Tanvee Gawand, Sahil Vaykar, and Neha Madhavi, of Computer Technology, Government Polytechnic Pen "JARVIS: An interpretation of AIML with integration of gTTS and Python " 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT).

[4]. Artur Dovbysh, Vladyslav Alieksieiev "Development and integration of speech recognition tools into software applications and an approach to improve of speech recognition quality " 2020 IEEE 15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET).

[5]. Ashok Kumar, Vikas Mittal "Speech Recognition: A Complete Perspective" International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7 Issue-6C, April 2019

[6]. Hardik Dhudrejia, Sanket Shah "Speech Recognition using Neural network" International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2278-0180, Volume-7 Issue-10, October 2018

[7]. Nisha "Voice Recognition Technique: A Review" International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653 Volume 5 Issue V, May 2017.

[8]. D. Snyder, D. Garcia-Romero, G. Sell, D. Povey, and S. Khudanpur, "X-vectors: Robust dnn embeddings for speaker recognition," in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018.

[9]. Satyam P. Todkar, Snehal S. Babar, Rudrendra U. Ambike, Prasad B. Suryakar Department of Computer Engineering Sinhgad College of Engineering Pune, India "Speaker Recognition Techniques: A review" 2018 3rd International Conference for Convergence in Technology (I2CT), Apr 06-08, 2018.

[10]. Garima Sood, Sidharth Pancholi, Amit M. Joshi Electronics and Communication Engineering Department National Institute of Technology, Hamirpur Himachal Pradesh, India "A Robust Speaker Identification System for Natural and Whispered Speech" 2017 IEEE.

[11]. Virtual assistant: What is it, 10 2017 onlineAvailable: www.searchcustomerexperiences.techtarget.com

[12]. V.Lalitha, A.Dinesh , L.Parameswaran, S.Dinesh Kumar, Department of CSE, Sri Sairam Engineering College, Chennai "ML Based Virtual Personal Assistant" International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 6, Issue 7, July 2019.

[13]. Prajyot Mane, Shubham Sonone, Nachiket Gaikwad and Prof. Jyoti Ramteke Computer Engineering Dept., Sardar Patel Institute of Technology, Mumbai, "Smart Personal Assistant using Machine Learning" International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS-2017).

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Fake Image and Document Detection using Machine Learning

Amit Lokre[1], Sangram Thorat[1], Pranali Patil[1], Chetan Gadekar[1], Yogesh Mali[1]
[1]Department of Computer Engineering , Dr D Y Patil School of Engineering, SPPU Pune, Maharashtra, India.

ABSTRACT

In the recent times, the rates of cyber crimes has been increasing tremendously. It has been proven incredibly easy to create fake documents with powerful photo editing softwares. Also social media has proven to be the largest producer of fake images as well. Various malpractices have also been on surge with the help of producing digitally manipulated fake documents. Detection of such fake documents has become mandatory and essential for unveiling of the documents/images based cyber crimes. The tampered images and documents will be detected using neural network .The output of the system will distinguish original document from a digitally morphed document. The system will be implemented using Neural Networks.

Keywords : Artificial Neural Network ; GLMC Features; Graphical User Interface ;Machine Learning ; Support Vector Machine.

I. INTRODUCTION

In the recent times the speed of cyber-crimes has been increasing tremendously it's been proven incredibly easy to make fake documents with powerful photo editing software Also social media has proven to be the most important producer of faux images also Various malpractices have also been on surge with the assistance of manufacturing digitally manipulated fake documents Detection of such fake documents has become mandatory and essential for unveiling of the documents/images based cyber-crimes The tampered images and documents are going to be detected using neural network The output of the system will distinguish original document from a digitally morphed document The system are going to be implemented using Neural Networks this is often an desktop application the rates of cyber-crimes are on a rise it's been proven incredibly easy to make fake documents with powerful photo editing software Documents and pictures are often scanned and morphed within minutes with the assistance of sort of software available On Investigation it States a foundation and it provides an answer to differentiate between original document and digitally morphed document Here the accuracy of system method has accuracy of 96 It is also possible to change metadata content making it unreliable here it's used as a supporting parameter for error level analysis decision . In, Xunyu Pan, Siwei Lyu proposed a scheme to detect the copy-move forgery in a picture , mainly by extracting the key points for extraction. The difference between the normal method and proposed scheme is first segments the test into semantically independent patches before key point extraction. within the second stage, to refine an estimated matrix an EM-based algorithm is employed and to verify the existence of copy-move forgery. The methods are categorized in two types as active

and passive forgery detection methods. The scope of system is restricted to review on passive forgery detection methods. System aims to present the study on different old methods of image forgery detection using different approaches like DWT (Discrete Wavelet Transform), SIFT, LBP (Local Binary Pattern). With the advancements in digital image acquisition, processing and reproduction technologies, the perfection has become easier with the fabrication of document. Document examiners examines the Composite copies which are produced for fraudulent purposes. The easy availability of those technologies to criminals leads to their application to preparation of fabricated photocopies or computer generated hard copies. Such document are produced in court of law because the only available piece of evidence with a excuse the the first document has been lost , eaten by moths or burnt during a fire etc. Superimposition is employed to detect counterfied documents, mainly tampered with employing a photocopier . This study presents examination of machine generated questioned document consisting of fabricated and manipulated writing signatures using digital image processing and reproduction tools. However, such techniques have now become obsolete since forgery lately is digital, clean and indistinguishable to the human eye. Therefore, machines are a more viable option now. Most of the techniques wont to detect those manipulations employ machine learning and pattern recognition. Image processing algorithms like DWT (Discrete Wavelet Transform) and SVD (Singular Value Decomposition) are one among the feature extraction methods that are used today to detect forged images Another devised scheme is to divide the image into overlapping blocks, thinking of them as vectors and find the manipulated region through radix sorting. Here method for detecting copy-move forgery over images tampered by copy-move. To detect such sort of forgeries, the input image is split into overlapping blocks of equal size, feature for every block is then extracted and represented as a vector, radix sort is employed to extract all the features of vector. Finally, the medium filtering and connected component analysis are performed on the tentative detected result to get the ultimate result. Compared with other methods, employing the radix sort makes the detection far more efficient without degradation of detection quality. Another approach to detect tampered images is to form use block based methods, but by using the non-overlapping texture blocks as a base for the graceful blocks, thus reducing the computational capacity. The algorithms then evaluates and compares supported their performances associated with a group of predefined parameters, this characterization are going to be used for further evaluation on the performance and efficient of an given blocked based cloning detection algorithms under the study. The result found after comparing them a user ready to select the foremost optimal forgery detection technique, counting on the user format and sort transformation it involves . Reflective SIFT based algorithms also are proficient in the detecting duplicated blocks in copy-move forgeries. due to this different detection techniques are suggests . we'd like to require care of image forgery. the pictures are often scaled, rotated, and flipped (mirror reflected). SIFT (Scale invariant feature transform) is usually wont to match images. But it fails for flipped images. Mirror-reflection invariant feature transform (MIFT) is presented in framework. Here, we'll find out how MIFT improvise SIFT

Literature survey:

Morphing images digitally has experienced tremendous growth in past 10 years Now a days several software are available that are wont to manipulate image in order that the image is appear as if as original Images are used as proof for authentication for any crime and if these images

aren't genuine then it'll be a problem for acceptance System is employed to detect these sorts of forgeries to work out whether a digital image is original or manipulated may be a big challenge to seek out the traces of tampering during a digital image may be a challenging task this technique presents a number of the Image Manipulation detection techniques like contrast enhancement detection splicing and composition detection image tampering etc Comparison of those techniques concludes the higher approach for its future research Using neural network tampered images are detected which recognizes the morphed region of the image and unveils the segments of the first image It are often implemented on Android platform The compression ratio of the changed content during a fake image is different and is detected using Error Level Analysis(ELA) that of the first image . Results prove the great performance of the scheme by comparing it with the state-of-the-art schemes on the general public database. the most two types are (a) Copy-Move forgery (b) Image splicing forgery of image Copy-move tampering is most generally employed by attackers during which object of another image is copied and pasted in original image in nearly matching areas. Hence to detect such image threats, it's required to automatic computer vision based method which may classify whether input digital image is original or tampered. there have been many methods introduced for copymove forgery detection from last 15 years.

II. METHODS AND MATERIAL:

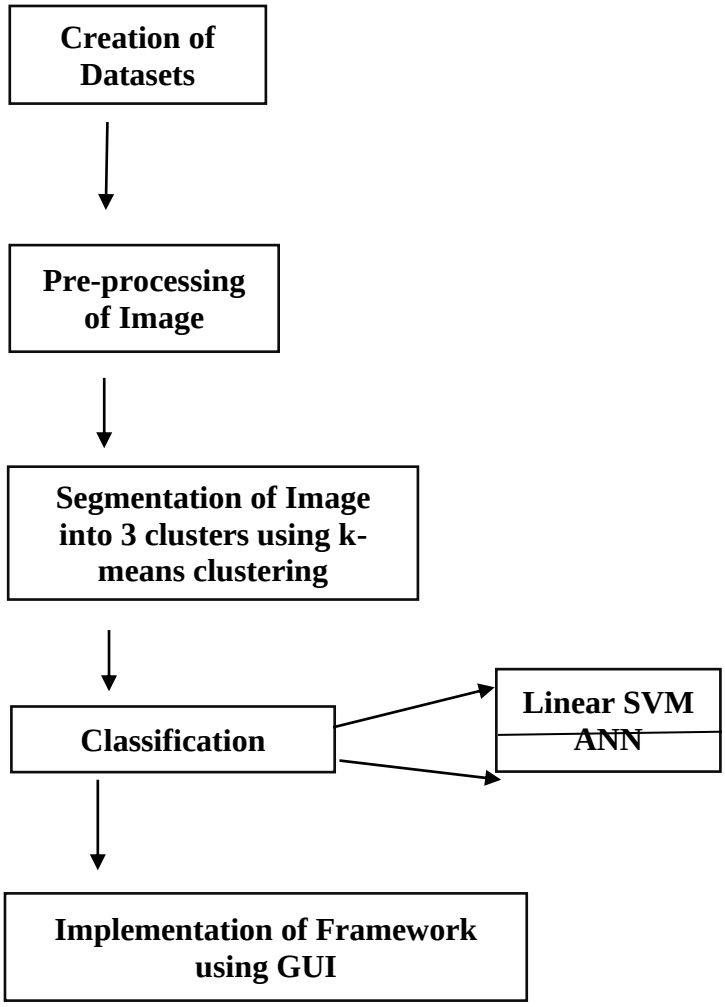


Fig 1.flowchart of methodology

1.Creation of dataset:

The images used for training this system are collected from various internet sources, college dataset and are morphed using photo editing tools. These images are edited using Adobe Photoshop CC 2017 to create a dataset of images- one original and its edited version.

2.Pre-Processing of Image:

To make the details of the images stand out more, the query image will be enhanced using histogram equalization .Histogram equalization: It is a necessary step because sometimes minute forgeries go undetected through the entire process. It is important

that the machine gets most of the details in one go. Histogram equalization, as the name suggests, is a method, where the intensities are adjusted using the histogram of the image. This technique is used here for contrast enhancement .Another essential stage in the pre-processing of an image is the removal of noise i.e. de-noising.

Denoise: De-noising is again done so that the details of the image are sharper and are not missed while extracting the features of the image.

Median Filter: A kernel of pixels is scanned over pixel matrix of the entire image. The median of a pixel values in the window is computed, and the center pixel of the window is replaced with the computed median. Since the median value must actually be in the value of one of the pixels in the neighborhood , the median filter does not create new unrealistic pixel values when the filter straddles an edge .For this reason median filter is much better at preserving sharp edges.

3.Segmentation:

Segmentation will be performed using k-means clustering, In this the image will be divided into K segments and the appropriate image will be chosen upon the data contained it .For this the GLMC feature will be used and the frame or segment having highest of the mean will be chosen. The GLMC frames will be compared to the original image and it will be helpful in calcutating the result as this value will be stored into some array and it will taken into count at the time of results.

4.Extraction of Features:

Feature extraction is a primitive type of pattern recognition and it is very important for pattern recognition.In this step we extracts some GLCM features such as Entropy, Angular Second Mo- ment, Contrast, Maximum Absolute Deviation, and Mean.Texture, shape and colour are features which are also extracted from an input image or document.Gray Level Co-occurrence Matrix is a tabulation of how often different combinations of pixel brightness values occur in a image. GLCM contains the information about the positions of pixel having similar gray level values.GLCM calculation units receive pairs of gray level values as input. The GLCM calculation unit consists the different combinations of gray values like a0b1, a2b3, a10b21 etc. This gives the deviation present in the image when compared with origi- nal image by predictive image.

5.Classification:

The interconnection of the network can be adjusted based on the number of available inputs and outputs making it ideal for a supervised learning.A linear kernel SVM will be one of the classifier used for classification and the other one will is ANN.

III. RESULTS AND DISCUSSION

This is the approach discussed in this paper executed successfully. The main purpose of the project was to inspect the handled computer legal documents for forensic testing using images. The processing technique that has been applied, the dataset required for the project was created by handling purchased images that no naked human eye could recognize it to handle image skill fully. After that, the feature information implemented by GALCM in MATLB, also provided by image processing toolbox R2015a examination and comparison required results between the original and the morphed legal document. Also HOG values were calculated but from the sort of matrix produced was very large trained in SVM so that they were eliminated.

TABLE I: Classifiers used and their results

PARAMETERS	ANN	SVM
Accuracy	96.4%	87.6%
Sensitivity	97.0%	89.0%
Specificity	95.0%	86.0%

Use of SVM, classification of differences between one between the above morphed and the original legal document the result would have been different meanwhile more property specific assets of two documents. Also, by other mathematics, the expression would have been more sophisticated consequences of splitting morphed documents more easily. The use of an ANN classifier provided high accuracy 96.4% less than linear SVM which gave less accuracy.

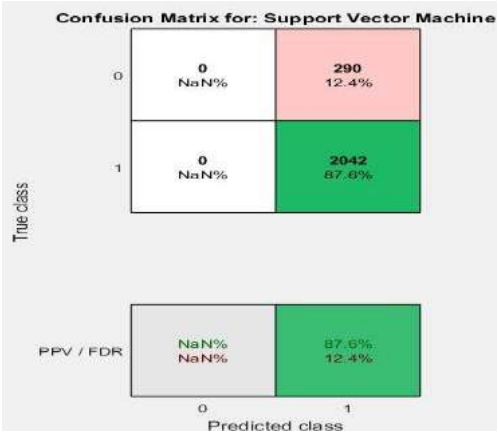


Fig 2: Confusion plot by Linear SVM

Fig 3: Confusion plot by ANN

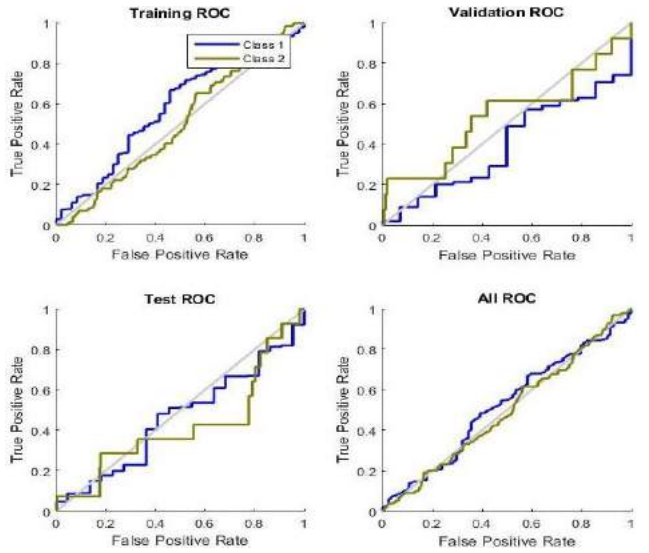


Fig 4.ROC plot by ANN

IV. CONCLUSION

Starting by editing images to crate a dataset and then its morphing and then scientifically examining the results obtained. Use the transformation technique above morphed and original image and finally edit the texture, features are the same initially, it was planned. At the present time, advances in science and technology, various advance introduction of image editing tools are also on the rise. These tools has multipurpose features. We can use this advanced image editing tools for the expansion of our next project execution to make results more quick and easy. While these tools are mostly used in areas related to creative design , criminals also can easily get access to it and as a result they can create a fake identity to hide themselves in a public place or the crime can be under investigation. These fake documents are not visible for the human eye. So, that was our intention purpose, such a method with good efficiency and accuracy will continue to refine procedures so that there are fewer loops in the analysis and hopefully this will come handy in the future.

V. REFERENCES

[1]. Shruti Ranjan, Prayati Garhwal, Anupama Bhan, Monika Arora, Anu Mehra,"Framework For Image Forgery Detection And Classification Using Machine Learning".2nd International Conference on Trends in. . . 2018. DOI: 10.1109/icoei.2018.8553924

[2]. Mohsen Zandi, Ahmad Mahmoudi- Aznaveh, Alireza Talebpour, "Iterative Copy-Move Forgery Detection Based on a New Interest Point Detector", Information Forensics and Security IEEE Transactions on, vol. 11, pp. 2499-2512, 2016, ISSN 15566013

[3]. Kushol, Rafsanjany Salekin, Md Sirajus Hasanul Kabir, Md Alam Khan, Ashraful. (2016). "Copy-Move Forgery Detection Using Colour Space and Moment InvariantsBased Features" .2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA 1-6.10.1109/DICTA.2016.7797027

[4]. Xunyu Pan, Siwei Lyu, "Region Duplication Detection Using Image Feature Matching", Information Forensics and Security IEEE Transactions on, vol. 5, pp. 857-867, 2010, ISSN 1556-6013.

[5]. Anil Dada Warbhe, Rajiv V. A Fast, Block Based, Copy-Move Forgery Detection Approach Using Image Gradient and Modified K Means"

[6]. Dharaskar, Vilas M. Thakare, "Digital image forensics: An affine transform robust copy-paste tampering detection", Intelligent Systems and Control (ISCO) 2016 10th International Conference on, pp. 1-5, 2016

[7]. Badal Soni, Pradip K. Das, Dalton Meitei Thounaojam. "(2018) CMFD: a detailed review of block based and key feature based techniques in image copy-move forgery detection". Institution of Engineering and Technology(IET) Image Processing 12:2, pages 167-178

[8]. In, Hwei-Jen Wang, Chun-Wei Kao, Yang-Ta. (2009). "Fast copy-move forgery detection".Article in WSEAS Transactions on Signal Processing".

[9]. Joshi MC, Kumar A, Thakur S. "Examination of digitally manipulated-machine generated document", a case study elucidating the issue of such unwanted progenies of modern technology". Prob Forensic Science 2011;56:162–73. Navoneel Chakrabarty, Sanket Biswas "A Statistical Approach to Adult Census Income Level

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Employee and Workspace Safety using WSN

Shreyas Lokhande[1], Narendra Choudhary[1], Aniket Chaudhary[1], Suraj Pethekar[1], Prof. Ajita Mahapadi[2]
[1]Department of Computer Engineering, Dr. D.Y.Patil School of Engineering, Lohegaon, Maharashtra India
[2]Assistant Professor Department of Computer Engineering, Dr. D.Y.Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

This paper provides an effective study of Wireless sensor network for air quality management for safety of user microcontrollers and sensors. It also discusses about the requirement of wireless nodes which keep monitoring of work place continuously for ensuring the security of both the employees as well as environment. Its advanced technology is already been applied in many fields. This paper focuses on the application of wireless sensors in various industries, and designs a set of security monitoring and alarm system using WSN network and GPRS network.

Keywords: WSN network, GPRS network

I. INTRODUCTION

Nowadays, the vast use of real applications and challenges favor the use of "Wireless Sensor Networks" primarily as they are more efficient and cost effective as compared to alternative approaches. This characteristic of Wireless Sensor Networks allows them the capability to deploy sensor vectors under various circumstances in military and civilian applications.

Many industries are part of a developing country. Like Petroleum and chemical industry as a very important pillar of the national industry, is an important support for the national economic take-off, but also is the country's key projects to achieve industrialization. Because of the increased demands of chemical and petroleum products in different fields, all kinds of dangerous products and high-pressure equipment must be installed and produced, so the safety of petrochemical plants is also very important.

WSN: Wireless sensor network (WSN) can be described as a group of dedicated and spatially dispersed sensors used for the purpose of monitoring and recording the physical conditions of the environment and then organizing the data which is collected at a central location. WSN measures different environmental conditions such as temperature, sound, pollution levels, humidity, wind, and so on. The Wireless Sensor Network consists of nodes which can range from a few to even thousands, where each node is connected to sensors. Each of such sensor network node consists several parts which includes a radio transceiver with an antenna (internal or external), a microcontroller, an electronic

circuit for the purpose of interfacing with the sensors and an energy source, which is generally a battery.

Data processing Issues require a massive processing capability, expertise in computing and a large amount of resources to perform computation and operations.

Parallel processing of data from WSN is effective way to handle this issue.

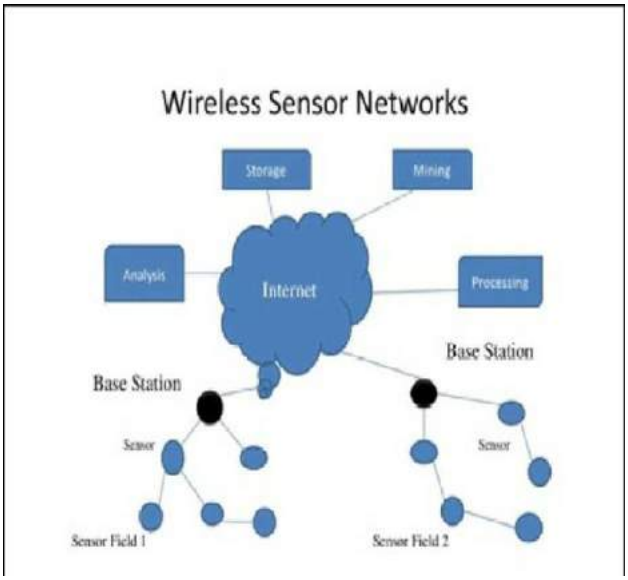


Fig 1. Wireless Sensor Network

II. GENERAL DESIGN

The structure of the alarm system and safety monitoring in petrochemical industries consists of two parts, first part is the internal network equipment, which includes security detection node, transceivers, upper computer, portable terminal management node (hand held); and the other part is the network equipment which includes a base station and GPRS remote terminal. The figure below shows the overall structure of the system.

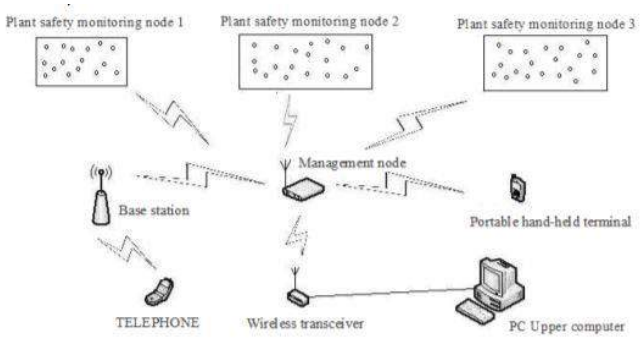


Fig 2. Overall Structure

The details of the danger information and environmental parameters can be gathered by querying the equipment, the combustible gas sensor immediately detects if the combustible gas exists in the field, sensor, then detects concentrations will be compared with the set value, if it is excessive, it is immediately reported to the concerned authorities, will measure the results sent to the upper computer at the same time and then show the location information and concentration size, and then the base station signal will be sent to devices in the form of textual messages.

When some other kinds of dangerous situations occur, the field nodes will be tested in similar ways. The position of the query and the specific parameters to be checked can be input by the key through the portable handheld terminal if the factory staff want to understand the situation, then the corresponding node will be real-time parameters sent to the portable terminal.

(1) Plant safety monitoring node: Plant safety monitoring node is similar to skin on human body which from time to time comes in contact with the external environment, it has the ability to perceive the information of the working environment of the plant area. It consists of various sensors such as humidity sensor, smoke sensor, temperature sensor, combustible gas sensor and camera, etc.

(2) Management node: The management node is mainly composed of wireless transceiver module, power module, GPRS module microcontroller.

(3) Portable hand-held terminal: It is used to check the field parameters and receive alarm information in the network.

(4) PC upper computer: The data command and wireless transceiver device can be connected to the data transmission by the RS232 serial port through the management software On the PC upper computer. The security detection and the alarm system interface of the upper computer is prepared based on the VC 6, insert MSCOMM serial communication control can achieve PC serial communication.

(5) GPRS remote terminal: GPRS remote monitoring is mainly refers to carry the mobile phone or any other connectable device, the environmental parameters of the scene can be queried and the alarm information can be received in any location through the phone or device.

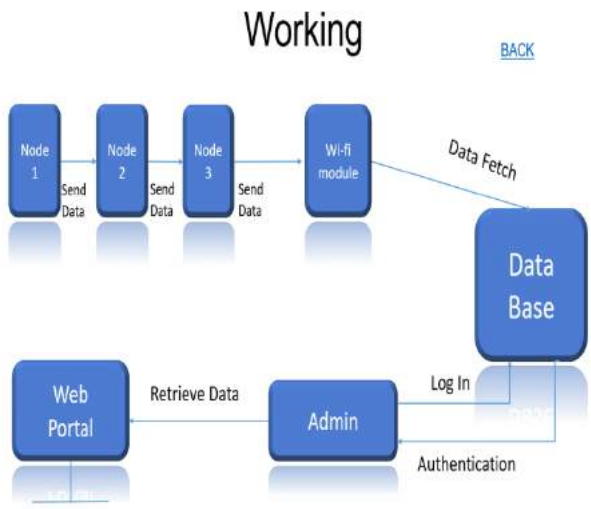


Fig 3. Basic Working module

III. POSSIBLE THREATS AND THEIR SECURITY MEASURES

One of the most dangerous attack of this category is the Sleep Deprivation Attack is in which intruders cause random drainage of sensor node batteries in order to shorten its lifetime dramatically. The network itself will be prolonged by detecting the SPA lifetime of a sensor nodes batteries. To compare values with predefined parameters to see if there are any intruders who are trying to harm the network and when found those malicious nodes are excluded Anomaly detection is used.

The DoS attack is another frequent attack, for the prevention of the DoS attack we introduce one interesting protocol. It should be mentioned that we have two types of the DoS attacks, nodes which uses network for its own purposes and communication (passive attacks) and harming other nodes unintentionally, and the malicious nodes that intentionally have a will to harm other nodes by not using the energy efficiently (active attacks). As it is in WSN, nodes need to forward messages to other nodes but in some cases, they cannot do that. This protocol acts as a game theory so that it is able to recognize those nodes that could act maliciously.

IV. CONCLUSION

Based on the alarm system and security detection application, it combined with the current management situation of various factories, designs a set of alarm system and security monitoring of plants, which includes portable handheld terminal, plant security detection node, GPRS remote terminal four parts and upper computer. The system has the function of preventing gas-leakage, fire prevention, and obtaining the spot monitoring parameters and image information at any time and at any time, it can

effectively improve the safety management of the plat in the plant and even prevents theft. By this use of technology, we can protect the factory employees and the safety of public property.

V. REFERENCES

[1]. Design of Safety Monitoring and Alarming System for Petrochemical Plant Based on WSN by Wang R and Zhu L.

[2]. Survey On Security In Wireless Sensor Network by Faris Fazlic and Seyed Ali Hashemi

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Voice Based Email for the Visually Impaired

Rahul Ahire[1], Poonam Bankar[1], Aniket Bhosale[1], Dipak Khette[1], Prof. Ajita Mahapadi[2]

[1]Department of Computer Engineering, Dr. D.Y.Patil School of Engineering, Lohegaon,Maharashtra India

[2]Professor, Department of Computer Engineering, Dr. D.Y.Patil School of Engineering, Lohegaon,Maharashtra India

ABSTRACT

E-mails are most reliable way of communication over Internet, for sending or receiving some important information. But there is a special criterion for humans to access the Internet and the criteria is you must be able to see.A survey shows that there are more than 285 million visually challenged people around the globe. That is, around 285 million people are unaware of how to use Internet or E-mail. So, forgiving an equal status to visually challenged people we have come up with this project idea which provides the client(user) with ability to send mails using voice commands without the need of keyboard or any other visual things.This system can be used effectively by handicapped and illiterate people as it is based on TTS, STT CONVERSIONS and IVR technologies.

Keywords: TTS, STT CONVERSIONS and IVR

I. INTRODUCTION

The Internet is a vast network which connects millions across the globe in various ways. So, talking about communication over the internet the first thing that comes to thought, is, E-mails. E-mails are extensively used form of online communication, both formally and informally as well. Despite social media, E-mails being the very traditional form of communication have still been the best to date. But the purpose of any service is to serve all mankind, and hence, E-mails should also be such that, they can be easily used by people from all races of life. But, Traditional E-mail Systems are accessible to several but the visually impaired class on the globe, and also various other handicapped people. So, in order to remove this drawback, An E-mail System for the visually impaired individuals would be an incredible breakthrough. Hence, this application has been thought of. Talking of the application, the application will be a web-based E-mail System for visually impaired people.Using Interactive voice response (IVR), which would enable everyone to control their own mail accounts using their voice only andalso they would be able to read, send, and perform all the other user tasks which areoffered by the traditional E-mail Systems. The system will prompt the user with voice commands to perform certain actions and the user will respond to the same with voice input. **The main advantage of this system is the use of (text) keyboard is completelyeliminated, which means, the user will have to respond through their own voice and mouse eventsonly.**Now you must be thinking that how will a impaired person willseetheright position on the screen for doing mouse clicks event. But this

system will perform actions based on the clicks only that is left or right click, it does not depends on the area(portion) of the screen where the cursor is placed before the click giving user the freedom to click anywhere on the screen.

1. SPEECH TO TEXT

Speech to Text conversion is the process of converting spoken words intotexts. This process is also called Speech recognition.

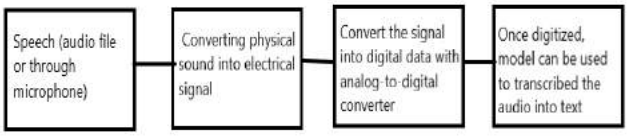


Fig 1.1 : Speech to Text Conversion Process.

2. TEXT TO SPEECH:

Text-to-speech (TTS) is also processthat lets your computer or phone read the text out aloud to you. Text-to-speech is commonly used as a feature to help people who have trouble reading the text from screen,but it is also convenient for those who want to be read to.

People with visual and reading inabilities were the early adopters of TTS(text to speech).

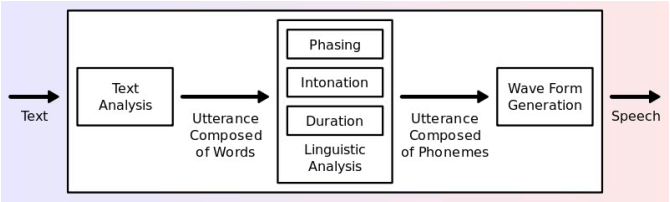


Fig 1.2 : Text to Speech Conversion Process.

II. OBJECTIVES

- To provide the visually impaired people, a Voice Based Mailing application, a platform using which they can easily receive or send emails without any third person's help or interference.
- Using Interactive Voice Response, people will control their mail accounts using their voice only and would be able to read, send, and perform all the other useful tasks.
- Use of keyboard is completely eradicated, the user will have to respond in the form of speech itself.

III. LITERATURE SURVEY:

Name of Paper	Author	Findings
Voice based E-mail system using AI	RijwanKhan , Pawan Kumar Sharma	Basic understanding of how the existing email system work and how we can resolve drawback of existing e-mail system using Artificial Intelligence
Voice based E-mail system	Prof. ManasiChoc he	huge number of people who cannot avail services of various other applications as well as email services due to they being visually impaired, so this project is being developed keeping in mind the hurdles faced by these people

IV. EXISTING SYSTEM

The Existing system does not support any voice commands or audio facilities and therefore it is not suitable for visually challenged people. Also,various existing search engine which take request in form of text from user and retrieve the relevant documents from the server and respond by displaying it in the form of text which is not accessibleby the visually challenged people.All operations in existing E-mail systems are dependent on mouse click events.

V. PROPOSED SYSTEM

The Proposed system will make the Traditional E-mail systems easily accessible to visually challenged people and also be of great help to the society. The Proposed system is being implemented, while keeping one idea in mind that it should be easily accessible for all kind of individuals.This system is accessible by any individual, whether they be visually challenged or not in an efficient manner.The ProposedSystem will take care of user-friendliness of traditional users, and will also focus on user-friendliness of all kinds of individuals.

In this system, the system is going to be prompting the user to perform specific operations to avail various services and if the user wishes to access various services then he/she has to perform that operation. Firstly, the user will have to register in the system through the registration form. The user goes to be assisted through voice commands, whereas while registering, all the mandatory fields to be filled are going to be scanned by the website; once the user would speak, it would get written automatically.After successfully registering, the user can log in by speaking the Username and Password when prompted by the system, this username and password will then be converted from speech to text and then the user will be authenticated by verifying the credentials with the database. Users can access various sections like Compose, Inbox, and Sent Mail after successful login.

VI. SCOPE OF THE PROJECT

- This type of system has great scope in the domain of HCI (Human Computer Interaction).
- The system would be beneficial to visually impaired people and it will help them to use modern day applications with ease.
- The Voice based technologies and NLP could be used in different applications as well such as gaming, robotics etc.

VII.ARCHITETURE DIAGRAM

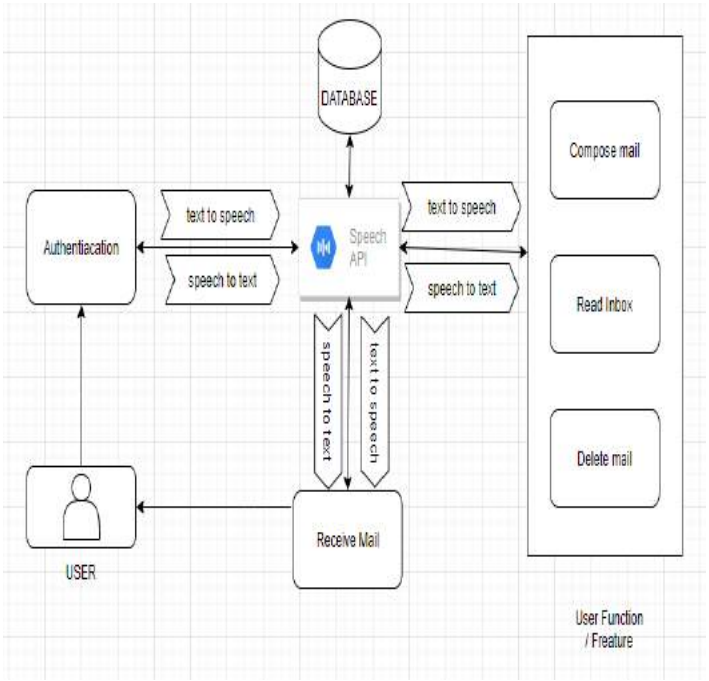


Fig 1.3 : System Architecture

VIII. CONCLUSION

This E-mail system can be used by any user of any age group having any physical inabilities with ease of

access. It has the features of speech to text as well as text to speech conversions with speech reader which makes designed system to be handled by visually impaired peopleconsiderably easy and efficient.

IX. REFERENCES

[1]. Suraj Tripathi1, Abhay Kumar1*, Abhiram Ramesh1*, Chirag Singh1*, Promod Yenigalla1, "Deep Learning based Emotion Recognition System Using Speech Features and Transcriptions". Samsung R&D Institute India – Bangalore, arXiv.org, 2019.

[2]. NithyaRoopa S., Prabhakaran M, Betty.P, Nov 2018. "Speech Emotion Recognition using Deep Learning". International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7 Issue-4S, November 2018

[3]. k Ashok Kumar, J L MazherIqbal "Machine Learning Based Emotion Recognition using Speech Signal". InternationalJournal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-1S5, December, 2019

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Helmet Detection on Two Wheeler Riders using Machine Learning and Automatic Licence Plate Recognition for Identification

Vaibhav Kharade[1], Rachana Jaykumar[1], More Pratik[1] , Mahendra Mahaja[1], Jayashree Chaudhari[1]
[1]Department of Computer Engineering, Dr. D.Y.Patil School of Engineering, Lohegaon,Maharashtra India

ABSTRACT

To ensure safety measures on road, detection of traffic offenders is a highly desirable but a very challenging task due to various difficulties such as closure, lighting, low video surveillance, various weather conditions, etc. this violation is a challenge due to the population and the low level of access caused mainly by the lack of an automatic system for detecting violations and taking necessary action. The growing number of people and the increasing number of vehicles make it impossible for manual systems to prevent this problem. The latest developments in Deep Learning and Image Processing provide an opportunity to solve this problem. This manuscript introduces the implementation of the three-component system which is a car, the non-use of a helmet and the number of the vehicle being monitored using Tensorflow. In-depth learning using SSD MobileNet V2 is the main method used to use the system. In this paper, we present a framework for automatic detection of motorcycle riders who drive barefoot in surveillance videos.

Keywords: SSD MobileNet V2, video surveillance, Dataset, Tensorflow and Deep Learning

I. INTRODUCTION

As motorcycles have become more affordable and have become the main route for many Indians. However, it comes with a high risk of accidents. To reduce the risk involved, it is advisable for motorcyclists to use a helmet. In view of this, the government has made it a criminal offense to ride a motorcycle without a helmet. Currently video-based methods are being used. This is simply not possible and is not possible as people are involved. The automation of this process is very desirable in recognizing the reliable and firmness of these violations on college premises. By looking at the properties and challenges required, we propose the idea of automatic acquisition of motorcycle riders without a helmet, using inputs from existing surveillance cameras operating in real time. Therefore, the solution to finding offenders using existing equipment makes it cost effective.

In order to ensure the safety measure, the detection of traffic rule violators is a highly desirable but challenging task due to various difficulties such as occlusion, illumination , poor quality of surveillance videos , varying whether condition ,etc.

This project holds an agenda to propose a framework for automatic detection of motorcyclists driving without helmets in surveillance videos.

Along with it we present a framework for Number Plate recognition as well. This will be achieved by use of Image Processing .

II. METHODS AND MATERIAL

fig.1 shows the structure of the SSD MobileNet The proposed project used a two-stage method provided by google, namely, the SSD MobilNet Version 2. The image database is divided into two parts for training and testing respectively.

Our goals will be achieved in four stages:
- Database collection.
- Annotation.
- Testing.
- Training assessment.

In machine learning, data collection is of great importance as it is a highly dependent category. We should be able to define all views and variations from potential data. Here the data can be collected by placing the camera on the road at the right angles to capture traffic.

The information is collected in bulk in open source videos available on any social media platform. After collecting images from all the resources can be divided into two categories namely training and testing. There are blank images added to the database to have a better knowledge base.

Models can be trained with images that contain a variety of objects, as well as a label that describes the category of the object it represents, and data that specifies where each item appears in the image. Next time the image is presented as a model entry, it will also release a list of items we found, state the location of the binding boxes containing each item, and the score indicating the level of confidence that the item has been found.

As the SSD model is trained to detect the presence and location of multiple categories of objects it leads to the release of four members included in the indices 0-4. All of these four inputs describe all the items in the input image, with one item in each list corresponding to each item in the image.

III. SYSTEM ARCHITECTURE

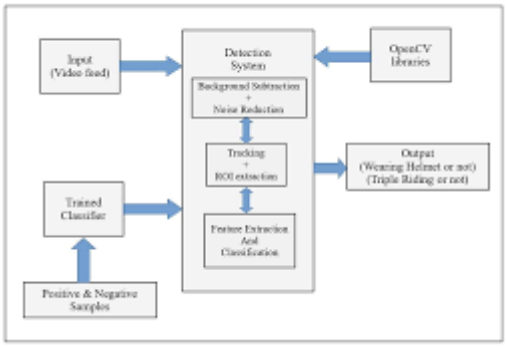


Fig 1

IV. RESULTS AND DISCUSSION

This is the complete flow diagram for our project starting from Giving video as an input to registering Number Plate Entry in the database.

The image (Fig.3) shows the Rectangle bounded objects and it can be clearly seen that the system is successfully able to detect all the desired objects while handling changes in conditions that are mentioned above. Model uses a matching phase while training, to match the appropriate anchor box with the bounding boxes of each ground truth object within an image. Basically, for the anchor box with the most significant level of cover with an article is responsible for foreseeing that item's class and its location.(Fig 3)

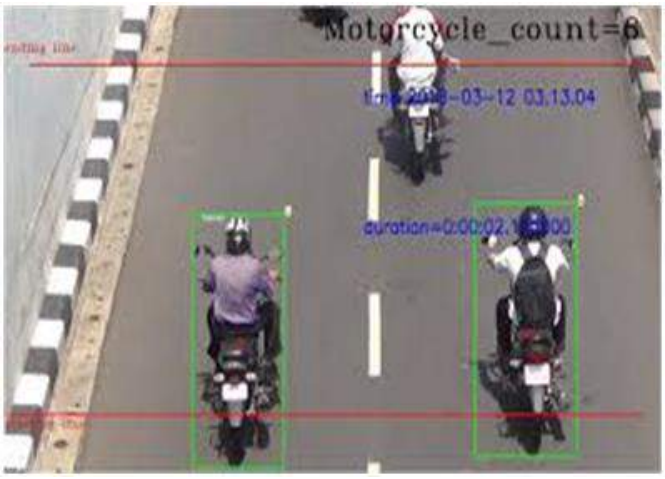


Fig 2

The model was prepared until the checkpoint for the 8,000 iterations were reached.. The final loss value at this checkpoint was 1.12. Table 1 represents the minimum, maximum, and average confidence score of the system under all the test conditions.

The variety in confidence is noted because of the adjustment in states of testing. This adjustment in the confidence factor under various conditions shows the versatility of the framework.The system performs well for all input cases.

The confidence can also be seen in the output images shown in Fig. 3. The images presented here are taken from the video processed using the system.

TABLE 1

Object name	Maximum, Average and Minimum Confidence scores seen in the video processed using the proposed system		
	Maximum confidence	Average confidence	Maximum confidence score.
BIKE	99%	87%	70%
NoHELMET	99%	81%	68%
NUMBER PLATE	99%	84%	71%

SSD MobileNet VS YOLO

Here we are contrasting two DNN object detection algorithms, a SSD (single shot detector) MobileNet and you-only-look-once (YOLO), through transfer learning.

The mobilenet-ssd model is a Single-Shot multibox Detection (SSD) network expected to perform object detection. This model is actualized utilizing the Caffe* framework.

You only look once (YOLO) is a state-of-the-art , real-time object detection system.

The recognition pace of the SSD MobileNet pith detector is better than that of the YOLO pith detector. It is equipped for making right pith detections 83.6% of the time.However, when the location error is thought of, the YOLO pith detector outperforms the SSD MobileNet pith detector since it has half the average location error. Both DNN pith detectors greatly outperform the non-DNN algorithm .

License Plate recognition

License plate recognition is one of the methods used for vehicle identification purposes. The sole purpose of this project is to find the most effective way to know the details of the License Plate from a digital photo (captured on camera). This process usually has three steps. The first step is License plate localization, regardless of the license-plate size and orientation. The next step is the segmentation of the characters and last step is the acknowledgment of the characters from the license plate. Thus, this project reveals the basic concept of the various algorithms required to achieve character recognition from a license plate during template simulation.

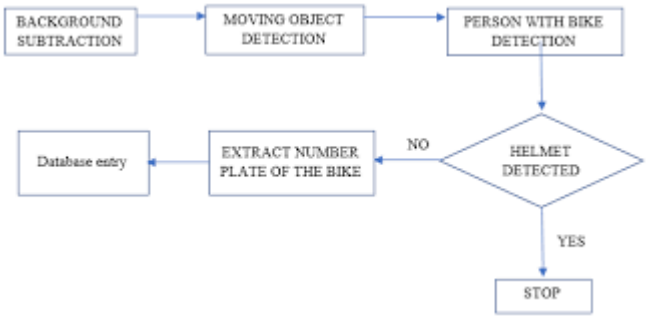


Fig 3

Steps followed:

Number plate recognition essentially comprises of three solid steps specifically:

1. Number Plate Extraction.

2. Character Segmentation.

3. Template Matching.

However, these steps are additionally divided into a progression of other steps whose working is as followed:

- Loading an RGB image
- Grayscale conversion
- Histogram equalization
- Binarization
- Dilation
- Edge detection
- Plate region extraction
- Character segmentation
- Template matching

All these steps can be carried out using the OpenCV.

And after successfully extracting the license plate we will put that entry in our database. Further necessary actions will be taken based on policies of authority.

V. CONCLUSION

The proposed program and project concludes by creating a solution that is able to identify features using transfer learning from custom databases. Therefore, the system demonstrates the possibility of achieving the goal of automating a traffic law enforcement process that enables additional functionality such as record keeping while completely eliminating manual input. The main focus of future additions is to increase the accuracy of the system by training it to suit a wide variety of vehicles and to add processing of individual number plates to create a database of recipients as compiled by the system. The whole system provides the need for a comprehensive monitoring and verification system to avoid positive false charges being added to the database.

VI. REFERENCES

[1]. Boonsirisumpun N, Puarungroj W and Wairotchanaphuttha P, Automatic Detector for Bikers with no Helmet using Deep Learning, Proc 22nd Int Comp Science & Eng Conf(IEEE)2018, 1–4.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Survey on Aid Donation Tracking Using Blockchain

Adesh Kolte, Prashant Chaudhari, Nihal Chhetri, Prof. Monika Dangore
Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

The Blockchain, a decentralized digital ledger system which has now been emerged as a technology that combines cryptographic, data management, networking, and incentive mechanism to support the verification, execution and recording of transactions between parties. The blockchain technology was first introduced for supporting the digital currency transactions in a secured way but now its potential has been recognized as a great importance in all form of transaction. Aid-donation taking and providing companies can take advantage of it by providing a transparent and safe transaction mechanism with smart contract system to the donors. It will allow the user to track their donation's and will get to know about the utilization of their donation hence makes this a crystal-clear transparent mechanism.

Keywords: Blockchain, Cryptographic, Data Management, Networking, Incentive Mechanism, Transactions, Digital currency.

I. INTRODUCTION

An increasing corruption and carelessness in different organization have led to decreasing faith of people towards these organizations. One of those organization are the aid-donation taking organizations which are meant to help others but is struggling from the same problem. Here blockchain can be introduced to create an online digital donation platform were every donor can donate easily and also they can track there donated amount and get to know about its utilization at ground level. This will increase the faith of people towards these organizations. Here user also have options to audit and also the smart contract system has been introduced here which is intended to automatically execute, control or document legally relevant events and actions as per the terms of an agreement. The basic objectives are to reduce the intermediations, arbitrations and enforcement costs, fraud losses as well as reduction of malicious and accidental exceptions.

II. METHODOLOGY

A. Proposed System:

In the existing system, the problem is. There is no track of the records of the money, transparency

- Payers are unaware of how their money has been utilized
- Fraudulent authorities indulge in corruption
- Corruption by the intermediaries administering funds and services.

These problems restrict payers from paying proper tax or donate in a campaign, as they are not sure about the legitimacy of authorities in the chain.

In the proposed system, the campaign creators will post their campaign for aid donation in the campaign and the interested people will donate the fund to the campaign. Where it defers from the old Aid donation platform is that all the money is now digital currencies like ether, dai. All ether coin will be recorded and keep tracks in the blockchain. Where the blockchain is an immutable ledger. The Donor has control over the funded money. The donor has full control over the money they have donated. By giving control on donated money the trust is built.

B. System Implementation:

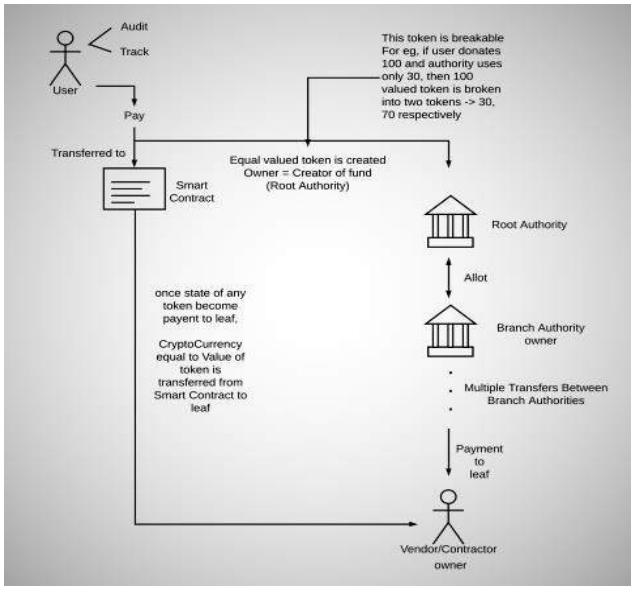


Figure 1: Flow of Aid Donation Tracking System

The proposed system is web application and smart contracts implemented using the solidity programming language. Solc is the solidity compiler used to compile the smart contracts and contracts file into bytecode and abi. The Bytecode will be deployed in the blockchain where the abi is in JSON format and is used to interact with the front-end. The front-end is designed using the React Js, Next Js, and Semantic-UI. The user interactive form will be used to donate easily. The creator or manager of the campaign or authority is to request money for donation. Moreover, the donor or payer can easily track; audit the funds that has donated to authority or campaign. Then authority provide donated amount to vendor. This will be recorded and stored in the blockchain.

C. Mathematical Model

Let's Assume input as a

{t}= token,

{N}: as amount of money (Cryptocurrency for eg: ether) user donates,

{T}: as a required amount allotted by authority to a vendor /contractor / owner

Input: {N}

{N} is amount in Cryptocurrency donated by donor

{T}= {N-t}

(The token is breakable if user donates {N} and authority uses only {N}-{t} amount {N} valued token is broken into two parts)

First Part: {T}={N}-{t}

Second Part: {t}={N}-{T} respectively.

D. Working Model and Architecture of project

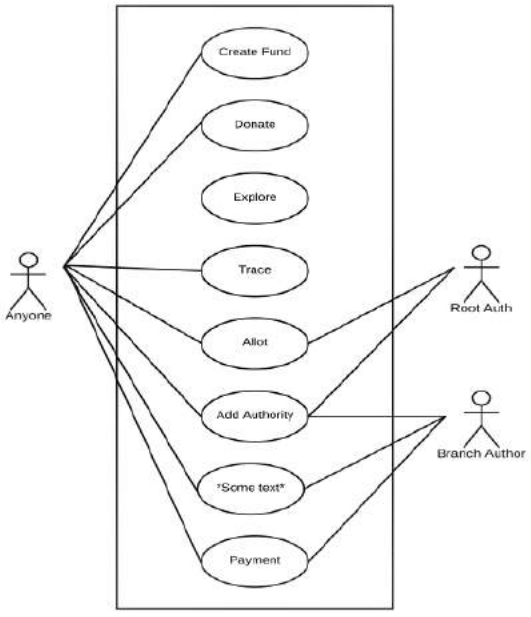


Figure 2: Working Model

A Both Aid Donation and Cryptocurrency is a trend on the internet and they match perfectly. Blockchain technology is one solution that can be used to reduce

the problems that occur in aid donation system. The contract is written in a way that all money will be sent to the authority or campaign. When the request meets the specified condition then required amount will be transferred to the Vendor. In addition, anyone can create funds add authority, donate amount to authority, trace the amount donated in authority and root authority can allot the donated amount to vendors.

Ethereum is an open-source, public; blockchain based distributed platform and operating to featuring smart contract functionality. It is the modified version of Bitcoin via transaction-based state transitions. Ether is a cryptocurrency that is generated and used by the ethereum platform. Ethereum provides a decentralized operating, the Ethereum Virtual Machine (EVM), which can execute an application on the public nodes.

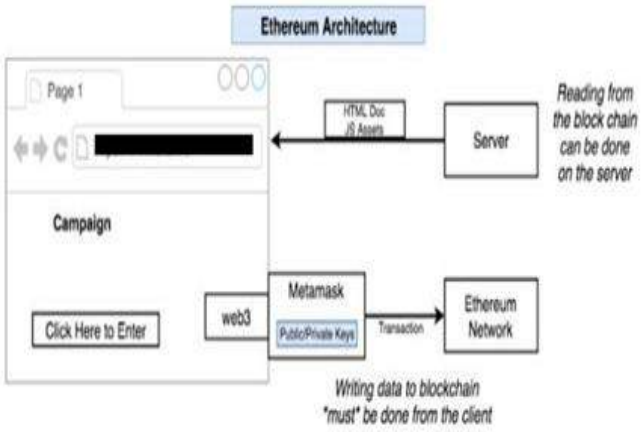


Figure 3: Architecture of Project

Figure 3 represents the architecture of the Ethereum, that shows how web application with solidity as backend works. Here server does far less work than what it does in a traditional web application. Server just sends some html, JavaScript content to the browser and but when user does an action or clicks on a button they do not reach server instead Ethereum application running inside the web browser uses web and communicates with Metamask, then Metamask creates a transaction signs it with user's private key and sends that transaction to Ethereum network. These transactions can be tracked by Ether scan using this procedure the aid donation system works.

III.CONCLUSION

Implementation of blockchain technology to Aid donation tracking provide transactions that are more transparent. As a result, users can feel more confident when they want to donate to an Authority. The application of smart contract on spending request also can help donor to know how their money are being spent. The purpose of a smart contract-based solution is to enable secure way of Aid donation by ensuring that the money donated by the donors is safe and also each and every step taken in the Authority with help of donated money has been tracked make a spending request where the purpose of using the money, to whom the money is being sent(vendor) and the amount needed should be mentioned. The main advantage of using the smart contract is the concept of blockchain that it is resilient against many threats. In addition, it provides many features like improved reliability, faster and efficient operation. It is user friendly, and has required options, which can be utilized by user to perform the desired operation. The expected goals can be achieved by the website are:

- Creating an Authority
- Contributing to Authority or campaign
- Track donated Fund
- Audit the donated Amount

IV. REFERENCES

[1]. Singh, A., Rajak, R., Mistry, H. and Raut, P., 2020, June. "Aid, Charity and donation tracking

system using blockchain.", In 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184) (pp. 457-462). IEEE.

[2]. de Vrij, Anna. "Blockchain in humanitarian aid: a way out of poverty and famine". Diss. Ph. D. thesis, 2018

[3]. Reinsberg, Bernhard. "Blockchain technology and the governance of foreign aid.", Journal of Institutional Economics 15, no. 3 (2019): 413-429.

[4]. Kumar, Randhir & Marchang, Ningrinla & Tripathi, Rakesh. (2020). "Distributed Off-Chain Storage of Patient Diagnostic Reports in Healthcare System Using IPFS and Blockchain." 1-5. 10.1109/COMSNETS48256.2020.9027313.

[5]. Saleh, Hadi & Avdoshin, Sergey & Dzhonov, Azamat. (2019). "Platform for Tracking Donations of Charitable Foundations Based on Blockchain Technology.", 182-187. 10.1109/APSSE47353.2019.00031.

[6]. Singh, Aashutosh & Rajak, Rohan & Mistry, Harsh & Raut, Prachi. (2020). "Aid, Charity and Donation Tracking System Using Blockchain." 457-462. 10.1109/ICOEI48184.2020.9143001

[7]. Nizamuddin, Nishara & Salah, Khaled & Azad, Muhammad & Arshad, Junaid & Habib ur Rehman, Muhammad. (2019). "Decentralized Document Version Control using Ethereum Blockchain and IPFS. Computers & Electrical Engineering." 76. 10.1016/j.compeleceng.2019.03.014.

[8]. Lin, Chao & He, Debiao & Huang, Xinyi & Khan, Khurram & Choo, Kim-Kwang Raymond. (2020). "DCAP: A Secure and Efficient Decentralized Conditional Anonymous Payment System Based on Blockchain. IEEE Transactions on Information Forensics and Security. PP." 1-1. 10.1109/TIFS.2020.2969565.

[9]. I. Sukhodolskiy and S. Zapechnikov, "A blockchain-based access control system for cloud storage," 2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), Moscow, 2018, pp. 1575-1578, doi: 10.1109/EIConRus.2018.8317400.

[10]. C. V. N. U. B. Murthy, M. L. Shri, S. Kadry and S. Lim, "Blockchain Based Cloud Computing: Architecture and Research Challenges," in IEEE Access, vol. 8, pp. 205190-205205, 2020, doi: 10.1109/ACCESS.2020.3036812.

[11]. M. D. Karumanchi, J. I. Sheeba and S. P. Devaneyan, "Cloud Based Supply Chain Management System Using Blockchain," 2019 4th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT), Mysuru, India, 2019, pp. 390-395, doi: 10.1109/ICEECCOT46775.2019.9114692.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Survey on Botnet and Its Detection Techniques

Shubham Gour[1], Yogesh Bhosle[1], Onkar Jagtap[1], Pratik Nirmale[1], Prof. Monika Dangore[1]
[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

Botnet term was coined when multiple networks of bots came into existence. It is a number of Internet-connected devices, which runs single or multiple bots. Botnets can be used to perform Distributed Denial-of-Service (DDoS) attacks, steal data, Ransomware, send spams, and allow attackers to gain unauthorised access on devices and its connections. Command and control(C&C) software are used by the Owner (BotMaster) to control the botnet. This paper explores the survey conducted on botnet and its detection techniques.

Keywords - Botnet, Botmaster,Intrusion Detection System(IDS), Neural Network, P2P, Network Traffic.

I. INTRODUCTION

A bot is an automated program which runs over the internet, some run automatically, while some run when they are triggered by specific input. Internet connected devices are infected with a piece of software that is bot. These internet connected devices are nothing but the botnet. After infection, these internet connected devices steer the instruction commanded by the owner of Botnet known as Bot Master/Bot Herder in 4 phases.

Following are the phases of the botnet infection:

Phase 1 Infection Initialization

A- "Social media" posts targeted by cybercriminal, In the first instance cybercriminal will post a malicious link on social media websites like hoax advertisements, shammed icons etc. When users perform any action on these websites, their action proved to be erroneous, as the current page is redirected to a malicious website, where the software gets installed which was already planted by the BotMaster.

B- "Infection method" approach is followed by the cybercriminals. In this "Email Phishing" tactics are being used to lure users on malicious websites as the user gets redirected when a link is being clicked, and their system gets compromised.

C- "Email Attachments" cybercriminals embody malicious pieces of software with an email, which gets downloaded once clicked and infects the whole system.

Phase 2 Connection to C2C Server

System manifests a connection with a command-and-control (C & C) server which establishes unauthorised connection periodically or may consummate upon infecting the system with malicious activity. Any infected machine liaising with C&C server will comply to launch a coordinated attack. e.g P2P, TELNET, IRC

Phase 3 Control

Cybercriminal (BotMaster) superintends the command and control of botnets for remote process execution by installing botnets on compromised machines. BotMasters uses Tor/shells to hide their tracks by hiding their identities via proxies to disguise their IP addresses.

Phase 4 Multiplication

Attacks in the first 3 phases are incessanted by Botmasters to infect copious internet devices by malicious use of botnet by promulgating fraud, spam emails, DDOS, keylogger, Miria botnet etc. Most recent attack was the " Kashmir Black", an active botnet comprehending thousands of compromised systems across 30 countries and exploiting dozens of vulnerabilities by targeting their CMS. It is believed that the campaign of the "Kashmir Black" started around the end of November 2019 and was trained to aim CMS platforms like Vbulletin, Opencart, Yeager, Joomla!, WordPress. Thus after knowing these vicious internet attacks which happen on a daily basis. We decided to counter this issue by implementing an ML model. In this paper we are going to fill up the canvass of loopholes and vulnerabilities with our ML model. To grasp the enormous nature of Machine Learning models let us first know about the basic model of Botnet. The Figure 1 stages a basic model of botnet in which botmaster is directly or indirectly connected to every other entity such as server, bots, benign hosts through two way communication.

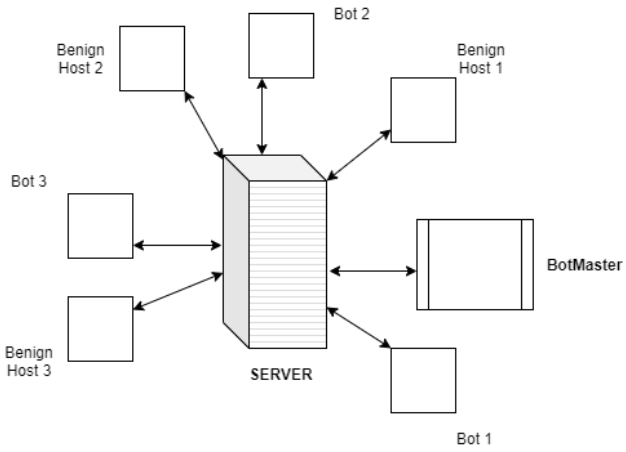


Figure 1.Model of Botnet

II. II.RELATED WORK

In the paper[1] author proposes botnet detection through graph-based feature clustering. In graph based feature clustering a novel botnet detection methodology has topological features like in degree, out degree, weights, clustering coefficient, node betweenness and centrality. A self methodize clustering is applied in networks to establish clusters of nodes based on these features. This technique is competent enough to isolate bots in clusters of small chunks while the majority of them are abided in the same big cluster, making them easily cast around a limited number of nodes. A procedure is developed to verify the algorithm efficiency by using[1] CTU-13 datasets against a detection method which is based on classification. Despite their assorted behaviour, results showed efficiency of the model is promising. In normal condition undesirable results may behave differently. Some models are trained to avoid the unfiltered data but the amount of humungus data challenges the computational expense.

The paper[2] proposes the detection of P2P botnets through network flow level behaviour analysis.In Network flow analysis an enhanced peer hunter

and Network flow based analysis system is used to detect peer to peer botnets. Once P2P network flow is detected then it moves to "correlative contacts" to cluster chunks of bots into communities. It uses community behaviour analysis to mark potential botnets. There are two evasion attacks, one where invaders know the attempts to evade the system by making peer to peer bots to act like legitimate P2P applications which showed high detection with some false positives can be achieved by using enhanced peer hunter technique. Botmaster uses three approaches adding randomized junk packets to depict bytes per characteristics , reducing the number of destination diversity rate from "mutual contacts", randomized spatial communication can be dealt if we adopt time communication features.

In the paper[3] the author proposes a system to detect botnets using a graph based behavioural model that extracts time dependent relations among the similar ones and detect the malicious connections. This process steers clear of the drawbacks due to statistical or signature-based techniques.

A graph is constructed based on the NetFlow records, after which an unsupervised mining method is used to trace root problem. Outlier detection and clustering techniques are used. From the flow records events are extracted from which sequence database is created using which graphs are plotted and outlier detection is used to detect anomalies.

The graphs generated are host-to-host and such graphs are generated for every individual host which entails a high overhead.

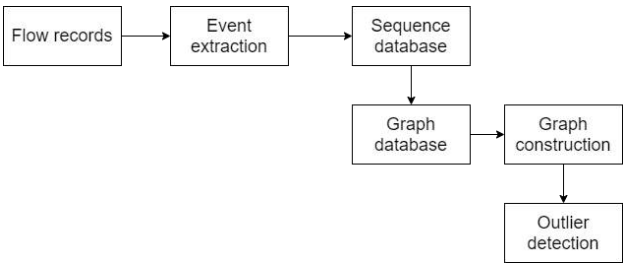


Figure 2.Steps in netflow records for BotGM

Genetic programming is used on streaming data by labels in this paper[4].Sometimes detection of a botnet behaviour is not as helpful, since bots can mix the malicious behaviour with normal behaviours or dynamically switch between different applications or using different versions of the same application. This model is being trained in a non-stationary stream of data. So, the prediction of a botnet cannot be apriori. The data can be labelled by humans to ensure the classification isn't manipulated by the attacker. The GP framework predicts the records that may be malicious, these records are then used to further train the classifier. In this method if the human operator who requests true labels, if he requests records from minor classes, then they are promoted aggressively which might result in reduced performance of major classes.

In the paper[5] Neural Networks are used for botnet detection.Botnets are one of the most increasing and significant threats to a large number of internet users in today's world. Botnet is a serious threat to the internet for committing cyber crimes such as DDoS attacks, stealing sensitive information etc.In this paper, the authors have proposed a Machine Learning technique for detecting P2P botnets. Convolutional Neural Network (CNN) has been used in the paper to train the model to detect the

botnets and Decision Tree for enhancing the decision rate to show the effectiveness of the proposed method for botnet detection. The proposed method transforms collected network traffic into flows, then extracts useful features and then utilizes algorithms to recognize botnets. Peer Rush and CTU datasets are used to validate robustness of the method proposed in the paper.

In the paper[6], indexing multiple for network traffic is used. In conventional firewalls, Intrusion Detection System (IDS) catalogues and manages network activities using the connection concept ‘‘flow', which encompasses five 'tuples' of source and destination IP, ports and protocol type. What is being fetched in a single connection can be obtained from TCP/IP field and packet content by inspecting the flows. In the era of the 21st century network has subsumed more connection and substantial bandwidth, the opprobrium of allowing permissions, granting access to IoT devices makes connection vulnerable. Triggering more malicious network threats, whose communications methods have been compromised. Factors like Miscalculations of duration of time and demand of instant data is responsible to acknowledge the network traffic behaviours to resolve this issue, additional two tuple and single tuple flow types being used to associate multiple 5 tuple communications, whereas discrete connection behaviour can be depicted by generating the metadata. These techniques ensure what network activities have been taking place over a course of time. To exemplify the proficiency of this approach a system rule set detects a Multi-peered Zeus botnet which communicates by establishing multiple connections with multiple hosts, thus imperceptible to standard IDS systems by monitoring 5 tuple flow types in isolation. As techniques are rule based, every parsing transpires in real time and does not require post-processing for further acknowledgment. This paper discusses the application for next generation firewalls and analyses the network traffic behaviour using those multiple tuple indexing.

In the paper[7] Machine learning has various applications and methods to solve real world problems in discrete domains. This is possible due to abundant data spread across over the network, significant furtherance of ML techniques, and advancement in computing capabilities. In the figure we discussed the components which are used to build a robust ML model for a given networking model.

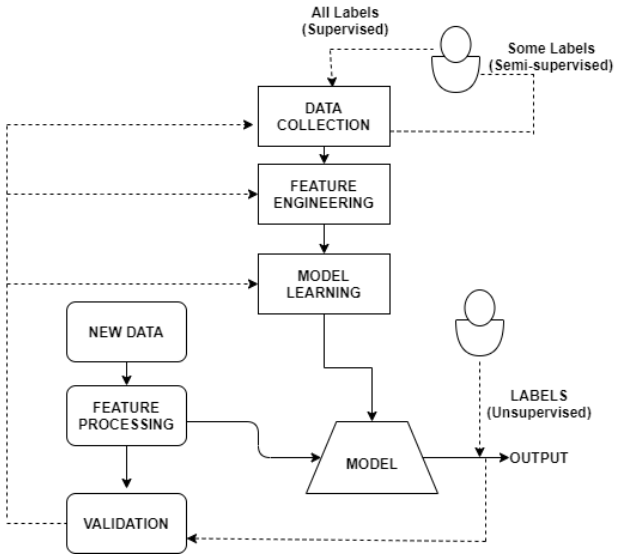


Figure 3. ML based solution.

ML has been applied to dispense its flexible nature to solve real world complex problems in network operations and other sectors. In survey we found out that perplexed problems across different network technologies can be unraveled by using diverse ML techniques which is an injunction with diverse application of Machine Learning.There are fragments like QOE , QOS

management, traffic prediction, congestion control, routing and classification management of networking which we have discussed in our paper to get the insights, scientific challenges and extent of ML in networking.Every effort is accountable and holds the responsibility to push the barriers of automatic network operations and their activities by using the features of ML in networking.

III. LITERATURE REVIEW SUMMARY TABLE:

Paper no.	Paper title	Methodology	Advantages	Drawbacks
1	Botnet detection using graph-based feature clustering	Topological features of nodes within a graph using clustering.	The flow-based detection systems have advantage over the packet-based systems.	If the dataset is large, the computational expense is often high for the detection approach, which is a huge disadvantage if faster detection is required
2	PeerHunter:Detecting peer-to-peer botnets through community behavior analysis	Flow based detection	Robust against changing behavior of botnets	The experiment results showed that the system is robust to PMMKL and will make the botnets less stealthy, less efficient and more exposed while conducting AMMKL But still there are chances that botnets are not detected.
3	BotGM: Unsupervised graph mining to detect botnets in traffic flows	Data mining	1.Uses netflow 2.Doesn't analyze individual behavior.	The graph generated for individual hosts entails a high overhead
4	On botnet detection with genetic programming under streaming data label budgets and class	Genetic programming	Works on a non-stationary stream of data.	If the minor classes are requested the efficiency decreases

	imbalance			
5	Botnet Detection Using Recurrent Variational Autoencoder	Convolutional Neural Network	Botnet detection using CNN has high accuracy rate and low false positive rate in detecting botnets as compared to other detection methods like anomaly based and signature based detection method	To conduct the classification of network traffic with the proposed technique is difficult as this technique is appropriate only for image and pattern recognition.
6	Peer Based Tracking using Multi-Tuple Indexing for Network Traffic	Multi-tuple indexing for network traffic analysis	Detection of multi-peered ZeuS botnet	Cannot locate features, generated by metadata to log traffic characteristics across multiple connections.
7	Survey on machine learning for networking: evolution, applications and research opportunities	Using ML in traffic prediction using MLP-NN	i.Traffic sampling and interpolation ii. Leveraging features other than traffic volume for traffic prediction.	It works on mostly on single-tenant and single-layer networks, re-architected is required to take into account the notion of multi tenancy in multi-layer networks.

IV. CONCLUSION

This paper construes various techniques and methods to deal with botnets under different situations over different networks. The main threat in bot detection is to avoid any loopholes or vulnerabilities in our own system while tracking them to terminate bot's network before their vicious goal is achieved by their botmaster.We realised that traditional techniques work fine but to make them further good and efficient we have to introduce Machine Learning . ML is the luminosity for the cataclysmic bots which exploit the vulnerabilities in networks and compromises the systems.

V. REFERENCES

[1]. Sudipta Chowdhury1*, Mojtaba Khanzadeh1, Ravi Akula1, Fangyan Zhang2, Song Zhang2, Hugh Medal1, Mohammad Marufuzzaman1, Linkan Bian1" Botnet detection using graph-based feature clustering".

[2]. Zhuang and J. M. Chang, "PeerHunter: Detecting peer-to-peer botnets through community behavior analysis,".

[3]. S. Lagraa, J. François, A. Lahmadi, M. Miner, C. Hammerschmidt and R. State, "BotGM: Unsupervised graph mining to detect botnets in traffic flows," 2017 1st Cyber Security in

Networking Conference (CSNet), Rio de Janeiro, 2017, pp. 1-8, doi: 10.1109/CSNET.2017.8241990.

[4]. Sara Khanchi, Ali Vahdat, Malcolm I. Heywood, A. Nur Zincir-Heywood,"On botnet detection with genetic programming under streaming data label budgets and class imbalance",Swarm and Evolutionary Computation,Volume 39,2018,,ISSN 2210-6502

[5]. Jeeyung Kim, Alex Sim, Jinoh Kim, Kesheng Wu," Botnet Detection Using Recurrent Variational Autoencoder".

[6]. Hagan, M., Kang, B., McLaughlin, K., & Sezer, S, "Peer Based Tracking using Multi-Tuple Indexing for Network Traffic".

[7]. Raouf Boutaba 1, Mohammad A. Salahuddin 1, Noura Limam 1, Sara Ayoubi 1, Nashid Shahriar 1, Felipe Estrada-Solano1,2 and Oscar M. Caicedo 2 "Survey on machine learning for networking: evolution, applications and research opportunities".

[8]. E. M. Hutchins, M. J. Cloppert, and R. M. Amin, "Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains," Inf. Warfare Security Res., vol. 1, no. 1, p. 80,2011.

[9]. S. Chen, Y. Chen and W. Tzeng, "Effective Botnet Detection Through Neural Networks on Convolutional Features," 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), New York, NY, 2018, pp. 372-378, doi: 10.1109/TrustCom/BigDataSE.2018.00062.

[10]. B. Alothman and P. Rattadilok, "Towards using transfer learning for Botnet Detection," 2017 12th International Conference for Internet Technology and Secured Transactions (ICITST), Cambridge, 2017, pp. 281-282, doi: 10.23919/ICITST.2017.8356400.

[11]. G. Vormayr, T. Zseby and J. Fabini, "Botnet Communication Patterns," in IEEE Communications Surveys & Tutorials, vol. 19, no. 4, pp. 2768-2796, Fourthquarter 2017, doi: 10.1109/COMST.2017.2749442.

[12]. H. Dhayal and J. Kumar, "Botnet and P2P Botnet Detection Strategies: A Review," 2018 International Conference on Communication and Signal Processing (ICCSP), Chennai, 2018, pp. 1077-1082, doi: 10.1109/ICCSP.2018.8524529.

[13]. C. Czosseck, G. Klein and F. Leder, "On the arms race around botnets - Setting up and taking down botnets," 2011 3rd International Conference on Cyber Conflict, Tallinn, 2011, pp. 1-14.

[14]. K. Alieyan, M. Anbar, A. Almomani, R. Abdullah and M. Alauthman, "Botnets Detecting Attack Based on DNS Features," 2018 International Arab Conference on Information Technology (ACIT), Werdanye, Lebanon, 2018, pp. 1-4, doi: 10.1109/ACIT.2018.8672582.

[15]. W. Zhang, Y. -J. Wang and X. -L. Wang, "A Survey of Defense against P2P Botnets," 2014 IEEE 12th International Conference on Dependable, Autonomic and Secure Computing, Dalian, 2014, pp. 97-102, doi: 10.1109/DASC.2014.26.

[16]. W. Sun and H. Gou, "The Botnet Defense and Control," 2011 International Conference of Information Technology, Computer Engineering and Management Sciences, Nanjing, Jiangsu, 2011, pp. 339-342, doi: 10.1109/ICM.2011.218.

[17]. M. Khosroshahy, M. K. Mehmet Ali and D. Qiu, "Scomf and SComI botnet models: The cases of initial unhindered botnet expansion," 2012 25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), Montreal, QC, 2012, pp. 1-5, doi: 10.1109/CCECE.2012.6334871.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Hand Sign Recognition Using Deep Learning based on Machine Learning

Darshan Ganatra[1], Omkar Shelke[1], Forum Makwana[1], Shivam Mishra[1], Prof. Nilesh Mali[1]
[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra, India

ABSTRACT

Addressing the issues of People with Hearing and Vocal Impairment through a single aiding system is a tough job. A lot of work in modern day research focuses on addressing the issues of one of the above challenges but not all. The work focuses on finding a unique technique based on the machine learning that aids the mute by letting them hear what is represented as text and its sound. The proposed system achieved the technique that takes the sign image through a web camera and applies to the image processing then analysis what exactly want to the mute people at end the text available to voice signals. All these three solutions were modulated to be in a single unique system. All these activities are coordinated using the Ubuntu system using python. The vocally impaired people are helped by the process in which the image to text and text to speech is given using machine learning.

Keywords: Image Processing, Tensor flow Algorithm, CNN Algorithm Process, Languages and compilers, Classification, Verification.

I. INTRODUCTION

Dumber people can simply tilt the message by sign language which could not be understandable by other people. By this system we provide the solution for blind, deaf and dumb people. For blind people the image is converted to voice by using Tesseract software, the deaf people received their content by message as soon as the opposite person speaks out it displayed as a message. The dumb persons conveyed their message through text instead of sign language which is delivered via e speak. We have provided necessary steps to resolve the problems of those masses.Approximately 285 million people are judged to be visually impaired worldwide in which 39 million are blind and 246 are said have low vision. Approximately 90% of this world's vocally impaired is from the dispirited income people and 82% of people living with blindness aging persons and above. The numbers of people visually impaired from eye related diseases have been brought down in the past 20 years according to global estimated work. In which 80% of all visual restitution can be prevented or cured. India is considered to be the home for the world's largest act of blind people. In this world, about 37 million are blind, in which 15 million are from India. There are so many researches have been getting along in this universe, but the visual impairment could not be broken for good. In lodge to facilitate these people we have developed the assistive device for blind people who does not want the assistance of other neighbours.

Goal and Objective:

- The great challenge lies in developing an economically feasible system so that physically impaired people can communicate easily.
- Datasheet of all the hand gestures will be made beforehand.
- Then, using python programing the real time picture of sign will be captured and will be compared with the datasheet. (photo captured will be converted into binary image).
- Then python will give the output based on Ubuntu system will be in accordance with the matched picture.
- At the end there is a sound that is being used to generate the voice message.
- There is a speaker through which message can be easily heard.

II. LITERATURE SURVEY

[1]. Vigneshwaran S, Shifafathimam, "Hand Gesture Recognition And Voice Conversion System For Dump People", in this paper author Explain two major techniques available to detect hand motion or gesture such as vision and non-vision technique and convert the detected information into voice through raspberry pi.
Advantages: Easy to wear hardware, Easy to design
Disadvantage: Sometimes fail hardware, Take time to detection

[2]. Trung-Hieu Le, Thanh-Hai Tran, Cuong Pham, "The Internet-of-Things based hand gestures using wearable sensors for human machine interaction" This paper introduces a new human hand gesture dataset which could be suitable for controlling home appliances. The dataset is captured with a low-cost and sensor plugable Internet of Things (IoT) device which is currently embedded with accelerometer and gyroscope sensors.
Advantages: Accelerometer and gyroscope sensors used, Easy to implement.
Disadvantage: Dataset used live predication not, sensor detection not work properly.

[3]. Rajit Nair Dileep Kumar Singh Ashu Shivam Yadav Sourabh Bakshi, "Hand Gesture Recognition system for physically challenged people using IoT", In this research a gesture-based recognition system has been developed that recognizes gestures with a Web camera in real time. The framework developed is based on a machine learning features and gestures. In order to improve efficiency and precision, we employ a number of steps in order to process and segment the picture before submitting it to the HGR Program. Our measures in image processing will identify movements in real time with high precision.

[4]. Vasileios Sideridis, Andrew Zacharakis, George Tzagkarakis, and Maria Papadopouli, "Gesture Keeper: Gesture Recognition for Controlling Devices in IoT Environments", This paper presents GestureKeeper which employs an accelerometer, gyroscope and magnetometer, from a wearable IMU, to first identify time-windows that contain a gesture, and then, recognize which specific gesture it is. GestureKeeper uses features based on statistical properties and acceleration samples. It can accurately recognize gestures from our 12- hand-gesture dictionary, exhibiting its best performance when the combination of features are used (e.g., about 96% mean accuracy). With the noise addition and feature selection, the mean accuracy is increased to over 97%.

III. PROPOSED SYSTEM

The great challenge lies in developing an economically feasible system so that physically impaired people can communicate easily.

- Datasheet of all the hand gestures will be made beforehand.
- Then, using python programing the real time picture of sign will be captured and will be compared with the datasheet. (photo captured will be converted into binary image)
- Then python will give the output based on Ubuntu system will be in accordance with the matched picture.
- At the end there is a sound that is being used to generate the voice message.
- There is a speaker through which message can be easily heard.

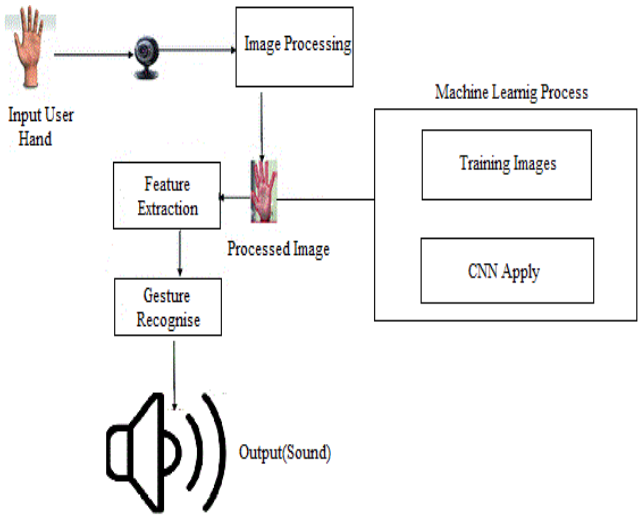


Fig1. System Block Diagram

Basic working algorithm of overall system is as follow

Algorithm basic Step:

1) Start
2) Load Hand sign Dataset for analysis.
3) Capture hand sign image from web camera.
4) Apply the image processing for covert the image gray scale.
5) CNN algorithm Match Gestures based on train dataset.
6) Show Text Output On Monitor
7) Output In Form of sound

IV. CONCLUSION

The implementation of the proposed system aims to translate gestures into speech (voice). The scope of the project is to enhance the recognition capability for various lightning conditions and achieving more accuracy. Implementing and identifying the more number of gestures. The miniature of the system should be done.

V. REFERENCES

[1]. Vigneshwaran S, Shifafathimam, "Hand Gesture Recognition And Voice Conversion System For Dump People", IEEE 2019.

[2]. Trung-Hieu Le, Thanh-Hai Tran, Cuong Pham, "The Internet-of-Things based hand gestures using wearable sensors for human machine interaction" IEEE 2019.

[3]. Rajit Nair Dileep Kumar Singh Ashu Shivam Yadav Sourabh Bakshi, "Hand Gesture Recognition system for physically challenged people using IoT" IEEE. 2020.

[4]. Vasileios Sideridis, Andrew Zacharakis, George Tzagkarakis, and Maria Papadopouli, "Gesture Keeper: Gesture Recognition for Controlling Devices in IoT Environments", IEEE 2019.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Smart Trolley With Advance Billing System

Niyamat Ujloomwale[1], Vaibhav Bandu Manwar2, Prince Kumar Singh[2], Patil Rohan Ranjit[2], Saurabh Shankar Ovhal[2]

[2]Assistant professor Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra, India

[2]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra, India

ABSTRACT

The shopping centre is a spot where individuals get their regular necessities. In shopping malls, there has been an emerging market for fast and simple payment of bills. Very sometimes, shoppers are dissatisfied with finding the items on the shopping list while shopping in a store and no help is required. We have developed a smart trolley with a smartphone app to solve these issues. This paper offers an interface to help consumers locate the product's location. It also offers a consolidated and automatic billing system using Node MCU's barcode scanner. Super markets will be issued with a barcode for each shopping mall commodity, to distinguish its type. A Product Identification System (PID) containing Node MCU, the barcode reader, is used for each shopping cart. Purchasing product details on the shopping cart can be read by a barcode reader and presented in the mobile app that is linked to the device. The complete bill is passed to the PC by the processor at the billing counter.

Keywords: billing trolley, barcode, nodeMCU, shopping.

I. INTRODUCTION

Since the dawn of civilization, people have continually produced creativity to support their demands. More independence can be the underlying explanation for success of creativity and this contributes to developing assignments and making them smaller and easier on a daily basis. Shopping is one crucial activity for individuals to expend the highest measure of energy. The shopping centre is a place where people get their everyday needs from food supplies, clothes, electrical equipment and so on. Most of the time clients have difficulties with the unspecific details concerning the object marked down and the abuse of the counters' superfluous time. Each grocery store and supermarket in this revolutionary world utilises shopping trolleys with a particular end purpose to help consumers pick and store the things they expect to purchase. Customers normally buy the necessary goods and put them in their carts and wait for bill payments at the counters thereafter. The paying of bills at the counters is a very troublesome and time-consuming procedure that raises the number of people at the counters. As demonstrated by a survey undertaken by the US Department Corporation, individuals regularly spend 1.4 hours shopping on a daily basis. If the queue is too long, a large amount of customers would choose to leave the

line. Two classifications (1) Shopping in-individual (2) Shopping in absentia can basically be defined as the new shopping environment. Shopping in absentia is maintained from multiple backgrounds, like web shopping, internet shopping, and so on, which would not entail physical keeping of the buyer at the counters. In-person buying requires an individual call at the place of purchasing and selecting products in view of various variables, including need, convenience, brand, and so on. The suggested keen shopping basket structure plans to assist person shopping to reduce the time spent shopping. In the normal time spent at the registers, continuous improvement is needed to boost the nature of the customers buying history. We also composed a shopping basket to solve these problems expressed above and to improve the present structure. This can be achieved by essentially adding barcode labels to the items in the shopping cart and the reader. Customers will provide details from this system relating to the cost of and thing that is within the cart and even the absolute cost of the item. With regards to the expense of the object, this structure would save time for shoppers and labour needed in the shopping centre.

A. Problem Statement

Realistically, markets are used by a large number of people these days to secure most products. The purchase of goods relates to an uncertain process requiring time spent on passageways, items and checkout lines. Consumers typically face some challenges and difficulties when shopping. Such issues entail stressing that the money they carried would be inadequate for all bought products and dissipating a lot of time at the cashier as well. And also because of the impact of disagreement and also because of lack of equipment that isolates application designs, it is becoming an increasing challenge for the merchants to keep their shoppers consigned and to foresee 3 of their demands. In some cases, consumers have concerns with the insufficient knowledge about the discount item and hence the misuse of superfluous time at the counters in some cases. With the barcode tag, we can end this problem. We present the systematic definition of barcode based keen shopping cart in the field of retail stock to address the current problems.

Objective:

The primary aim of this initiative is to introduce a smart shopping cart to improvise ordering with the aid of barcode technology. In the buying cart, the aim is to use the barcode-related monitoring deployment practise. The barcode is used in this plan to secure the purchase of commodities in shopping malls. If the product has been put in the shopping cart, the price of the product shows and the total amount is displayed accordingly, then if we want to delete the product from the trolley, the product will be taken away and the value of the particular product is deducted from the total amount. In this the equipment used is to purchase the goods, thereby increasing the efficiency and speed of protection when buying in shopping complexes. The technical goal of our presented issue in shopping complexes is the practise of barcode technology for the instinctive identification of goods within the shopping cart, thereby annihilating the shopper engaged in the buying and payment task of commodities. The key point of the suggested system is to include an invention that is minimally effort-oriented, easily adaptable, and efficiently viable to facilitate individual shopping. A lot of time can be saved at the billing counters with the help of this .

II. LITRATURE SURVEY

A shopping market is a place where customers come to purchase their daily using products and pay for that.

So there is a need to calculate how many products sold and generate the bill for the customer, when we go to shopping mart for shopping, we have to work for selecting the right product. Also, after that, it is hectic to stand in line for billing all the goods. Hence, we are proposing to develop a smart electronic trolley for shopping malls that keeps the track of purchased products and also helps the customer to pay the bill with the help of swiping machine provided in a trolley.so that the customer can save the time [1].

With the increasing employment of broad area wireless sensor networks in the field of consumer applications, it becomes imparitive to address the concerns raised by its application, such as reliability, energy consumption and cost-effectiveness. They described the implementation of a reliable, fair and cost efficient smart shopping cart using wireless sensor networks which is suitable for supermarkets that can reduce the man power. Here, we are using wireless sensor networks such as GSM to send the information read by the barcode reader to the main server of the shop and to the customers mobile[2]. Microcontroller based design has acquired the status of most happening field in electronics. This is specialized field that has power of integrating thousands of transistors on a single silicon chip. They proposed a microcontroller based trolley which contains barcode scanner. Customer has to place the product wrapper in front of the barcode scanner. The corresponding information will be displayed on the display. At billing counter customer has to pay and collect the bill. This smart eletronic trolley contains a barcode scanner and also a swiping machine on it, so that there is no need to go to the billing counter and pay the bill, instead customer can pay the bill by swiping his card. The total bill will be sent to the customers mobile, so that paper usage for billing can be reduced[3]. Intelligent shopping cart was being developed to assist aperson in everyday shopping in terms of reduced time spent while purchasing.The developed system consists of three components (a) server communication component which contains connection with the cart and the main server, (b) user interface and display component, (c) automatic billing component[4]. The smart electronic trolley contains a cart which is fully enclosed and there is opening and closing systems to reduce the steeling of products. By this there will be no worry of the fting for the shop processor in the shopping mall. With a wide spectrum of demands from a multitude of spaces, the present world has an increasingly growing population. Customers who need to purchase multiple items in shopping centres or grocery stores need a lot of time and patience to negotiate with each other for successful shopping and need to use our technology wisely to solve these issues. In the field of shopping, there are various new up-coming advances that offer efficient and cost-effective solutions. A few methods that have been suggested so far in genius shopping are discussed in this chapter. Aryan Prateek et.al. The Smart Shopping Cart with Automated Billing System via RFID and Bluetooth was suggested. A Product Identification Package (PID) containing the Bluetooth module, EEPROM, RFID scanner, LCD and microcontroller [5] is inserted in each shopping basket within this system.

Gangwal Udita et.al. Proposed Wireless Sensor Networks Smart Shopping Cart for Automatic Billing Purpose They portray the use of a secure, reasonable and cost-productive smart shopping cart using remote sensor networks in this article. For shops, such a system is ideal where it can support in-store laboratories and customers[6].

Sudhir Rupanagudi Rao et.al. A Novel Cost-Effective Savvy Trolley System based on Video Processing was proposed for supermarkets using FPGA. This depicts a new realistic technique to defeat the dilemma of being unable to find objects by attaching a web camera to the trolley[7].

A. Using GSM and adhoc wireless routing, Vijayaraj and R.Saravanan suggested Automatic EB Billing System Using GSM and adhoc wireless routing, they propose a mechanism in which power charging turns out to be entirely robotized and communication is conceivable through remote systems[8].

Suganya.R, and.al. Automated Smart Trolley using Arduino and Smart Billing. This device is based on RFID and Bluetooth [9] identification.

Radio-Frequency Identification, as stated in[10] Aboli Hanwate and Poonam Thakare (2015), is a technology that uses radio waves to pass data from an electronic tag, called an RFID tag or mark, attached to an object, through a reader to identify and monitor the object. A special type of wireless card that has incorporated the embedded chip together with the loop antenna is the RFID Sticker. The 12 digit card number is represented by the inbuilt embedded chip. The RFID reader is a circuit that produces a magnetic signal of 125KHZ. The loop antenna associated with this circuit, which is used to read the RFID card number, transmits this magnetic signal. The RFID card is being used as a security access card in this project.

III. BLOCK DIAGRAM

A. Barcode reader :

The barcode reader (or barcode scanner) is an optical scanner which is capable of reading printed barcodes, interpreting barcode data, and transmitting data to a device. It consists of a light source, a lens and a light sensor, like a flatbed scanner, which converts optical impulses into electrical signals. In addition, almost all barcode readers provide decoder circuits that can interpret the image data given by the sensor in the barcode and transfer the content of the barcode to the output port of the scanner.

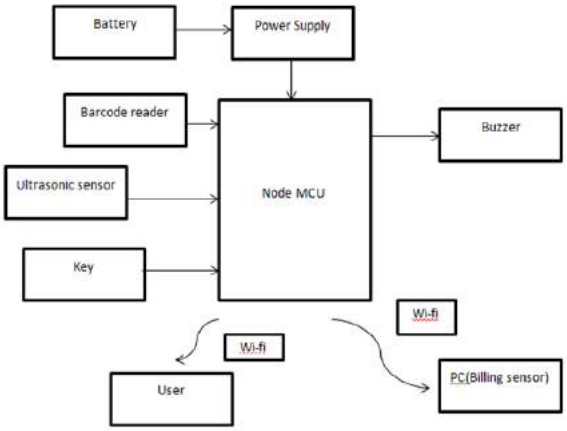


Fig.no.1.Block diagram of design system

B. NodeMCU:

The production board for the NodeMCU ESP8266 comes with an ESP-12E package with an ESP8266 chip and a Tensilica Xtensa 32-bit LX106 RISC microprocessor. This microprocessor supports RTOS and runs with a clock frequency adjustable from 80MHz to 160 MHz. For data and applications, NodeMCU has 128 KB of RAM and 4 MB of Flash memory. It is suitable for IoT ventures due to its high computing capacity and in-built Wi-Fi / Bluetooth and Deep Sleep Operating functions.

The Micro USB jack and VIN pin (External Supply Pin) can be used to power the NodeMCU. It supports the interface between UART, SPI, and I2C.

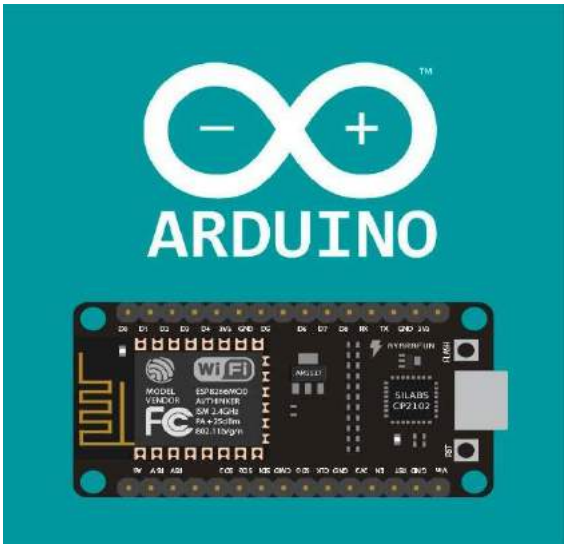


Fig.no.2. NodeMCU ESP8266

A. 12v battery:-

Choose one that suits the criteria until you have measured the total current (I). For eg, for the 600mA power supply, a 1Amp one. Now for our power transformer T, the rms secondary voltage (primary is whatever is compatible with your area) must be our desired output Vo PLUS, the voltage drops through diodes (two diodes). Vo1 must be adequate to provide the LM7805's minimum operating input requirements at all times (min 7.3v to max 25v). assume it has a central tap for a 24vac secondary output like the O.P.'s one (12vac from each terminal to the central tap, 24vac via terminals), so here the secondary output is taken from the central tap to one of the terminals. If it doesn't have a central tap and all you have is 24vac performance, due to the high Vp for this particular situation, it is advised to choose a different one.

IV. CONCLUSION

In the built prototype model, the desired goals were successfully achieved. The product produced is convenient to use and economical. While the project shows the proof of principle, to make the smart shopping cart more stable, there are a few things that should be added. To begin with the latency of wireless communication with the server will need to be taken into account in this project. Secondly, it is not really safe to communicate. Sticking the barcode sticker on certain items is difficult. Conventional barcode scanning is more advanced in such situations. In comparison, a more advanced micro-controller and a larger display system can be used to offer a greater interface for users.

V. REFERENCES

[1]. A. Farahzadi, P. Shams, J. Rezazadeh, and R. Farahbakhsh, “Middleware technologies for cloud of things-a survey,” Digital Communications and Networks, Elsevier, 2017.

[2]. D. Singh, G. Tripathi, and A. J. Jara, “A survey of internet-ofthings: Future vision, architecture, challenges and services,” in 2014 IEEE World Forum on Internet of Things (WF-IoT), March 2014, pp. 287–292.

[3]. A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, “Internet of things: A survey on enabling technologies, protocols, and applications,” IEEE Communications Surveys Tutorials, vol. 17, no. 4, pp. 2347–2376, Fourthquarter 2015.

[4]. A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, “Internet of things for smart cities,” IEEE Internet of Things Journal, vol. 1, no. 1, pp. 22–32, Feb 2014.

[5]. J. Rezazadeh, M. Moradi, A. S. Ismail, and E. Dutkiewicz, “Superior path planning mechanism for mobile beaconassisted localization in wireless sensor networks,” Sensors Journal, IEEE, vol. 14, pp. 3052–3064, 2014.

[6]. M. Hubert, M. Blut, C. Brock, C. Backhaus, and T. Eberhardt, “Acceptance of smart phone-based mobile shopping: Mobile benefits, customer characteristics, perceived risks, and the Impact of application context,” Psychology and Marketing, vol. 34, no. 2, pp. 175–194, 2017.

[7]. J. Rezazadeh, M. Moradi, and A. S. Ismail, “Efficient localization via middle-node cooperation in wireless sensor networks,” in International Conference on Electrical, Control And Computer Engineering, June 2011, pp. 410–415.

[8]. M. Moradi, J. Rezazadeh, and A. S. Ismail, “A reverse localization scheme for underwater

acoustic sensor networks," Sensors, vol. 12, pp. 4352–4380, 2012.

[9]. R. Nallanthighal and V. Chinta, "Improved grid-scan localization algorithm for wireless sensor networks," Journal of Engineering, Hindawi, vol. 5, no. 10, pp. 21–27, 2014.

[10]. P. Martin, B.-J. Ho, N. Grupen, S. Mu˜noz, and M. Srivastava, "An ibeacon primer for indoor localization: Demo abstract," in Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings, 2014, pp. 190–191.

[11]. J. Rezazadeh, M. Moradi, A. S. Ismail, and E. Dutkiewicz, "Impact of static trajectories on localization in wireless sensor networks," Wirel. Netw., vol. 21, no. 3, pp. 809–827, 2015.

[12]. J. Wang, P. Urriza, Y. Han, and D. Cabric, "Weighted centroid localization algorithm: Theoretical analysis and distributed implementation," IEEE Transactions on Wireless Communications, vol. 10, no. 10, pp. 3403–3413, 2014.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Emotion Detection to Prevent Suicide

Tejashri Sawant[1], Manorama Shewale[1], Supriya Kiwade[1], Amruta Chitari[1]
[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

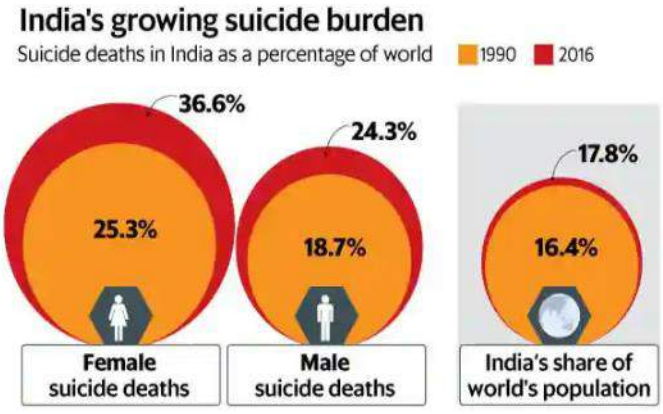
Suicide might be considered as one of the most serious social health problems in the modern society. Suicidal ideation or suicidal thoughts are people's thoughts of committing suicide. It can be regarded as a risk indicator of suicide. India is among the top countries among in the world to have annual suicide rate. Objective of Face Emotion Recognition (FER) is identifying emotions of a human for reduce the suicide rate. This system involves extraction of facial features, and threshold detection of stress using emotions expressed through face using the Convolutional Neural Network (CNN) algorithm. This system is basically used to classify positive and negative emotions and detects the stress based on usual threshold value.

Keywords: Suicide rate, Emotions, Convolutional Neural Network.

I. INTRODUCTION

Suicide is an important issue in the Indian context. More than one lakh (one hundred thousand) lives are lost every year to suicide in our country. In the last two decades, the suicide rate has increased from 7.9 to 10.3 per 100,000. There is a wide variation in the suicide rates within the country. The southern states of Kerala, Karnataka, Andhra Pradesh and Tamil Nadu have a suicide rate of > 15 while in the Northern States of Punjab, Uttar Pradesh, Bihar and Jammu and Kashmir, the suicide rate is < 3. This variable pattern has been stable for the last twenty years. Higher literacy, a better reporting system, lower external aggression, higher socioeconomic status and higher expectations are the possible explanations for the higher suicide rates in the southern states.

In 2016 the number of suicides in India had increased to 230,314. Suicide was the most common cause of death in both the age groups of 15-29 years and 15-39 years. About 800,000 people die by suicide worldwide every year, of these **135,000** (17%) are residents of India, a nation with 17.5% of world population.



India reported an average 381 deaths by suicide daily in 2019, totalling 1,39,123 fatalities over the year,

according to the latest National Crime Records Bureau (NCRB) data. A 3.4 per cent increase was observed in suicides during 2019 (1,39,123 suicides) as compared to 2018 (1,34,516) and 2017 (1,29,887), the data showed. The rate of suicide (incidents per 1 lakh population) rose by 0.2 per cent in 2019 over 2018, as per the data. According to the statistics by the NCRB, which functions under the Union Home Ministry, the suicide rate in cities (13.9 per cent) was higher as compared to all-India suicide rate (10.4 per cent) in 2019.

Suicide by ''hanging'' (53.6 per cent), "consuming poison'' (25.8 per cent), ''drowning'' (5.2 per cent) and ''self-immolation'' (3.8 per cent) were the prominent means of committing suicides during the year, the data showed.

India had the highest suicide rate in the South-East Asian region in 2016, a new report by the World Health Organization (WHO) has revealed. India's own official statistics, which map the number and causes of suicides in the country, have not been made public for the last three years, hindering suicide prevention strategies and efforts to execute the WHO's recommendations in this regard.

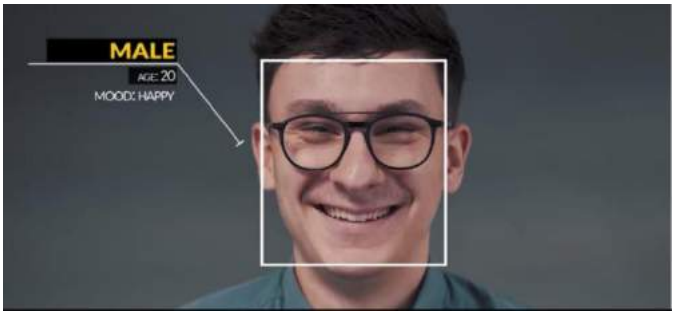
India's suicide rate stood at 16.5 suicides per 100,000 people in 2016, according to the WHO report. This was higher than the global suicide rate of 10.5.

The report presented suicide rates for countries and regions using data from the WHO Global Health Estimates for 2016. When classified according to region and income, India is part of the South-East Asia region and the Lower Middle-Income group of countries. India's suicide rate (16.5) was higher than the rate of its geographic region (13.4) and the rate of its income group (11.4).

II. METHODS AND MATERIAL

1. Face Detection

Face Detection is the first and essential step for **processing**, and it is used to detect **faces** in the images. A **facial detection** system uses biometrics to map facial features from a photograph or video. It compares the information with a database of known **faces** to find a match. **Face detection** systems use computer algorithms to pick out specific, distinctive details about a person's **face**.



These details, such as distance between the eyes or shape of the chin, are then converted into a mathematical representation and compared to data on other **faces** collected in a **face** database.

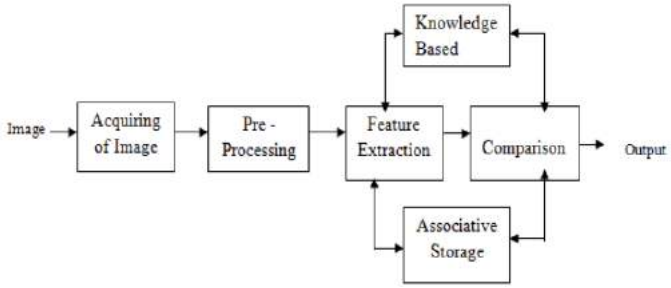
2. Emotion Detection

Emotion detection is used to analyse basic facial expression of human. **Emotion recognition** system is constructed, including face **detection**, feature extraction and facial expression classification.

© 2020 IJSRST | Print ISSN: 2395-6011 | Online ISSN: 2395-602X | Published : 18/12/2020 - 142-145

3. Methodology

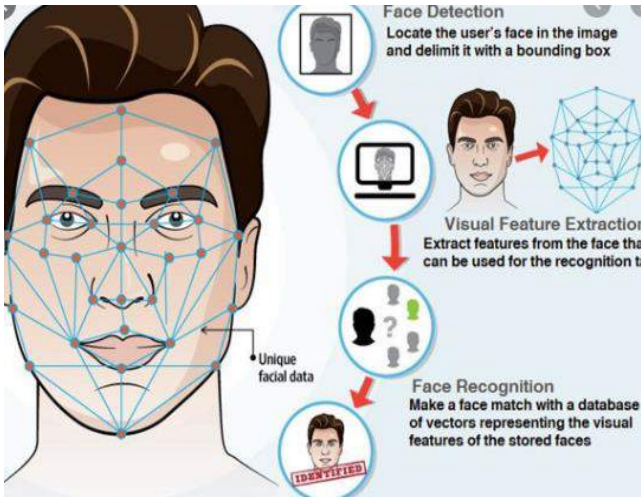
Face of the subject is captured using the camera module. This detected face is processed and the emotions are classified as either positive or negative emotions. The detected image is processed to identify the face of the subject using Convolutional Neural Network (CNN) algorithm



This is plotted and an increase in the negative emotion can be inferred as increase in stress

4. Feature Extraction

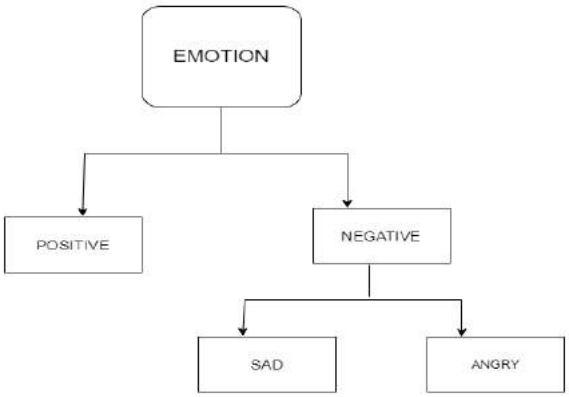
Facial feature extraction is the process of **extracting face** component **features** like eyes, nose, mouth, etc from human **face** image.



Facial feature extraction is very much important for the initialization of processing techniques like **face** tracking, **facial** expression recognition or **face** recognition.

5. Emotion Recognition

The emotions are to be extracted from the detected face. The image that is captured from the camera module, contains the facial features. The detected face is pre-processed (i.e.) cropped and resized. The detectors defined prior can be utilized to identify the emotion and sort them. It must be noted that viola-jones algorithm uses adaboost algorithm with cascading classifier, wherein a series of weak classifier's classification with a satisfactory threshold is combined to give an acceptable outcome.



6. Set Stress Threshold:

After the emotion is detected, it is plotted against the time axis. Over a period of time it is observed and if it crosses an estimated threshold value, some action is performed.

III.CONCLUSION

The predictor is relatively successful at predicting test data from the same dataset used to train the classifiers. However, the predictor is consistently poor at detecting the expression associated with contempt. This is likely due to a combination of lacking training and test images that clearly exhibit contempt, poor pre-training labeling of data, and the intrinsic difficulty at identifying contempt. The classifier is also not successful at

predicting emotions for test data that have expressions that do not clearly belong exclusively to one of the seven basic expressions, as it has not been trained for other expressions. Future work should entail improving the robustness of the classifiers by adding more training images from different datasets, investigating more accurate detection methods that still maintain computational efficiency, and considering the classification of more nuanced and sophisticated expressions.

IV. REFERENCES

[1]. Ekman, P. & Keltner, D. (1997). Universal facial expressions of emotion: An old controversy and new findings. In Segerstråle, U. C. & Molnár, P. (Eds.), Nonverbal communication: Where nature meets culture (pp. 27-46). Mahwah, NJ: Lawrence Erlbaum Associates.

[2]. Rajiv Radhakrishnan, Chittaranjan, "Suicide: An Indian perspective", Indian Journal of Psychiatry, (2012).

[3]. Soman C, Vijayakumar K, Ajayan K, Safraj S, Kutty V, "Suicide in South India: a community-based study in Kerala", Indian J Psychiatry, (2009), Vol.51, pp.261-264.

[4]. Deb, Esben Strodl, Jiandong Sun, "Academic Stress, Parental Pressure, Anxiety and Mental Health among Indian High School Students", International Journal of Psychology and Behavioral Science, (2015), Vol.5, Issue.1, pp.26-34.

[5]. P. Viola and M. Jones, "Rapid object detection was using a boosted cascade of simple features", CVPR, (2001), pp.511–518.

[6]. Damir Filko, Goran Martinovi´c, "Emotion Recognition System by a Neural Network Based Facial Expression Analysis", Automatika,(2013), Vol.54, Issue.2, pp 263–272.

[7]. Neha Gupta1 and Navneet Kaur, "Design and Implementation of Emotion Recognition System by Using Matlab", IJERA, (2013), Vol.3, Issue 4, pp.2002-2006.

[8]. Seyedehsamaneh Shojaeilangari, Wei-Yun Yau, Karthik Nandakumar, Li Jun, and Eam Khwang Teoh, "Robust Representation and Recognition of Facial Emotions Using Extreme Sparse Learning", IEEE Transactions on Image Processing, (2015), Vol.24, No.7, pp.2140-2153.

[9]. Bosker, Bianca, "AFFECTIVA's Emotion Recognition Tech: When Machines Know what you're feeling", (2013).

[10]. Vikramjit Mitra1etal, "Cross-Corpus Depression Prediction From Speech", ICASSP, (2015), pp.4769-4773.

[11]. Fuji Ren, Xin Kang, and Changqin Quan "Examining Accumulated Emotional Traits in Suicide Blogs with an Emotion Topic Model" IEEE Journal of Biomedical And Health Informatics, (2016), Vol.20, Issue.5, pp.1348-1351

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Online E-Voting System using Blockchain Technology

Shubham Kumar[1], Abhishek Patil[1], Geeta Kotwani[1], Sharan Patil[1], Prof. Chaitanya Bhosale[1] Prof. Prashant Mandale[1]

[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

India is the world's largest democracy with a population of more than 1 billion; India has an electorate of more than 668 million and covers 543 parliamentary constituencies. Voting is the bridge between the governed and government. The last few years have brought a renewed focus on to the technology used in the voting process. The current voting system has many security holes, and it is difficult to prove even simple security properties about them. A voting system that can be proven correct has many concerns. There are some reasons for a government to use electronic systems are to increase elections activities and to reduce the elections expenses. Still there is some scope of work in electronic voting system because there is no way of identification by the electronic voting system whether the user is authentic or not and securing electronic voting machine from miscreants. The proposed system is to develop a compatible voting machine with high security by using Block-chain technology in order to increase security and transparency between the users.

Keywords:- Electronic Voting System, Voter ID, Security, Block Chain, Vote

I. INTRODUCTION

Voting, whether traditional ballet based or electronic voting (e-voting), is what modern democracies are built upon. In recent years' voter apathy has been increasing, especially among the younger computer/tech savvy generation. E-voting is pushed as a potential solution to attract young voters. For a robust e-voting scheme, a number of functional and security requirements are specified including transparency, accuracy, auditability, system and data integrity, secrecy/privacy, availability, and distribution of authority. Block-chain technology is supported by a distributed network consisting of a large number of interconnected nodes. Each of these nodes have their own copy of the distributed ledger that contains the full history of all transactions the network has processed. There is no single authority that controls the network. If the majority of the nodes agree, they accept a transaction. This network allows users to remain anonymous. A basic analysis of the block-chain technology suggests that it is a suitable basis for e-voting and moreover, it could have the potential to make e-voting more acceptable and reliable.

II. Related Work

This paper [1], proposed secure voting system with fast voting results through RFID based biometric voting system. In this paper, there are two

verification steps involved. First, RFID tag is used which contains the verification data which is already stored in LPC 2148. Second, the Fingerprint scanner is used to check whether the RFID is belonging that particular person or not. The drawback of this paper is cost maximized due to use of RFID method.

In this paper [2], used of Aadhaar card provided by UIDAI with QR code present in it. Online instead of offline mode and storing the voting data to secured online server. Results can be displayed by admin after entering user id and password.

The proposed method [3] is to build a Smart voting system using fingerprint recognition technology that allows any voter in INDIA to cast the vote to their respective constituency from anywhere in INDIA by going to their nearest voting booth in the place of stay. Also to develop a secure smart voting system based on biometric recognition. Provides the voter to vote from any region with in India to their Residential Constituency from the nearest Voting Booth with a secure voting process without neglecting to vote.

This paper [4], proposes protected voting system to avoid the unlawful voting. The authentication of an individual is made using biometric and capability of the voter is affirmed using the Aadhaar. In this system the data stored in the Aadhaar card act main criteria for authentication and conformation. The security is provided through biometrics such as fingerprint. The fingerprint information stored in the Aadhaar is taken as the reference and used for authentication at the time of voting.

Basic electronic machine [5] which is used nowadays has some laggings like multiple vote casting from one member and invalidity of votes are checked automatically. To reduce these disadvantages, the smart automatically processed and fingerprints are used to reduce multiple vote casting in simple way.

This paper [6] has shown the possibility of establishing E-Voting protocol based on public-key encryption cryptosystem. The security of the proposed E-Voting depends on RSA public key encryption protocol. It allows the voter to vote from his/her own personal computer (PC) without any extra cost and effort. This protocol is proposed to replace the unreliable previous voting system, since voters feel justifiably confident that their votes will be counted.

This system [7] provides security from all type of attacks, when vote is travelling from voting client to voting server from their experimentation. These attacks include security threats from passive as well as active intruder. For authentication of voter instead of USERNAME, if we can use thumb impression of voter or capture photo of his/her face and compare it with photo stored in our database, it will be more secure.

In this paper [8], a block-chain-based voting system. It needs time to popularize block-chain for a voting system as it is a novel idea and voting itself is a crucial matter in a democratic country.

The proposed [9] model is more secure than other models and it is suitable for use in major elections on a large scale. After casting a vote with NCVVS system, the voter receives a confirmation email containing the ballot fingerprint (and also the fingerprint of the election) calculated by standard hash function SHA (256) [46].

The proposed work [10] is based on the block-chain technology, which remove all the threats from the communication link. It is a decentralized system, contain hashing and encryption concept for providing the security.

III. Existing approach

A lot of work has been done in this field thanks to its extensive use and applications. This section mentions some of the approaches that have been implemented to achieve the same purpose. These works are mainly

differentiated from the algorithm for E-voting systems.

The existing machine had security risks that can potentially undermine the election process. In addition to human error; internet e-voting is susceptible to a range of threats such as hacking by domestic and foreign saboteurs, technical glitches, voter impersonation and even system failure.

IV. Proposed Approach

In our system Block Chain Concepts are applied to Online Voting System when we are developing a Smart E-voting system by taking advantage of block Chain concepts with web interface.

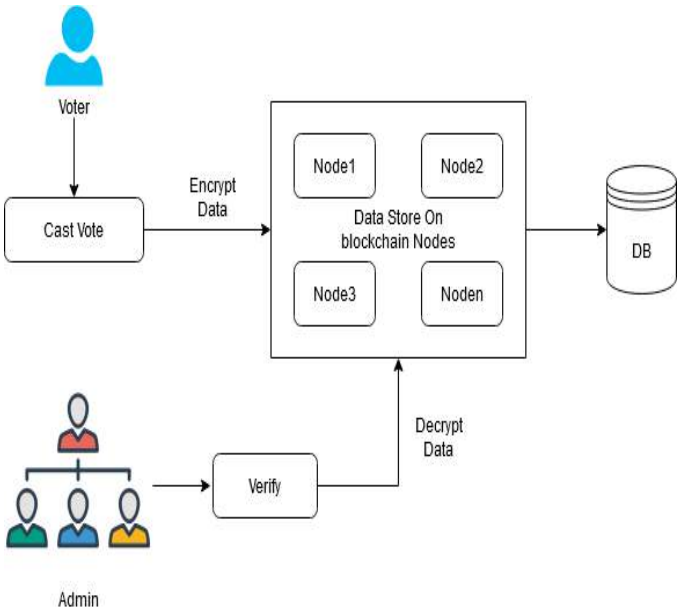


Fig 1. System Architecture

V. CONCLUSION

This paper described, an electronic Voting system for small to medium sized Internet-based public opinion systems that provides privacy of vote, voter's authentication, auditability, security, double-voting prevention, fairness voting device from manipulating the authenticated voters voting choices.

VI. Acknowledgement

Special thanks to Dr. Pankaj Agarkar, HOD , Prof. Chaitanya Bhosale And Prof. Prashant Mandale, Department of Computer Engineering, Dr. D Y Patil School of Engineering for guidance and resource provision which helped in development of this project. Thanks to all associated faculties for providing all necessary help.

VII. REFERENCES

[1]. J.Deepika, S.Kalaiselvi, S.Mahalakshmi, S.Agnes Shifani, "Smart Electronic Voting System Based On Biometric Identification-Survey", International Conference on Science Technology Engineering Management (ICONSTEM).

[2]. Ravindra Mishra, Shildarshi Bagde, Tushar Sukhdeve, J. Shelke, "Review on Aadhaar Based Voting System using Biometric Scanner", International Research Journal of Engineering and Technology(IRJET).

[3]. Girish H S, Gowtham R, Harsha K N, Manjunatha B, "Smart Voting System", International Research Journal of Engineering and Technology (IRJET).

[4]. K. Lakshmi, R. Karthikamani, N. Divya "Aadhar Card based smart e-voting system", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 8958, Volume-8, Issue-2S, December 2018.

[5]. G.Saranya, R.Mahalakshmi, J.Ramprabu, "Smart Electronic Voting Machine surveillance", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 8958, Volume-8, Issue- 2S, December 2018.

[6]. Ashish Singh, Kakali Chatterjee, SecEVS: Secure Electronic Voting System Using Blockchain Technology, International Conference on Computing, Power and Communication

Technologies (GUCON) Galgotias University, Greater Noida, UP, India. Sep 28-29, 2018.

[7]. Cosmas Krisna Adiputra, Rikard Hjort, and Hiroyuki Sato, A Proposal of Blockchain-based Electronic Voting System, Second World Conference on Smart Trends in Systems, Security and Sustainability.

[8]. Jena Catherine Bel.D, Savithra.K, Divya.M, A Secure Approach for E-Voting Using Encryption and Digital Signature, International Journal of Engineering Development and Research.

[9]. Abhijit J. Patankar, Kotrappa Sirbi, Kshama V. Kulhalli, "Preservation of Privacy using Multidimensional K-Anonymity Method for Non Relational Data", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-2S10, September 2019.

[10]. Ashraf Darwish and Maged M El-Gendy, A New Cryptographic Voting Verifiable Scheme for E-Voting System Based on Bit Commitment and Blind Signature, International Journal of Swarm Intelligence and Evolutionary Computation.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

A Framework for Analyzing Real-Time Tweets to Detect Terrorist Activities

Akshay Karale[1], Pranav Shinde[1], Pushpak Patil[1], Sanjay Parmar[1], Prof. Niyamat Ujloomwale[2]
[1,2]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra, India

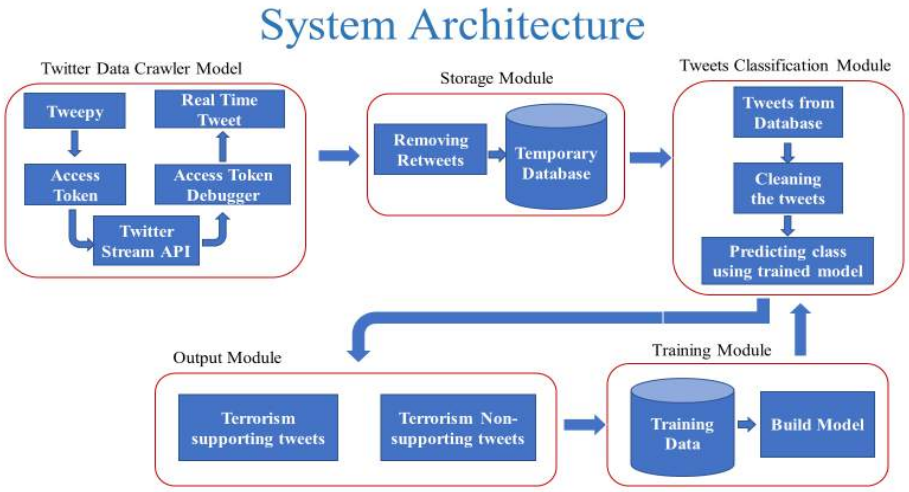
ABSTRACT

Terrorist organizations use different social media as a kind of tool for spreading their views and influence general people to join their terrorist activities. Twitter is one of the the most common and easy way to reach mass people within short time span. We have focused on the development of a system that can automatically detect terrorism-supporting tweets by real-time analyzation. In this system, we have developed a front-end system for real-time viewing of the tweets from twitter that are detected using this system. We also have compared the performance of the two different machine learning classifiers, Support Vector Machine (SVM) and Multinomial Logistic Regression and found that the first one works better than the second one. As our system is highly dependent on the data, for more accuracy we added a re-train module. By using this module wrongly classified tweets can be added to the training dataset and can train the whole system again for better performance. This system will help to ban the terrorist accounts from twitter so that they can't promote terrorism, their views or spread fear among general people in society.

Keywords: - Social Media, Twitter, Terrorism, Real-Time Tweets, Machine Learning.

Problem Statement: A Framework for Analyzing Real-Time Tweets to Detect Terrorist Activities.

The function of twitter data crawler module is to crawl real-time tweets from Twitter using Twitter Streaming API. The storage module stores the tweets temporarily. Tweet classification module predicts the category of the tweet. The output module shows the output of the system. The training module builds the classification model which is used to predict the category of each tweet.



© 2020 IJSRST | Print ISSN: 2395-6011 | Online ISSN: 2395-602X | Published : 18/12/2020 - 150-155

I. INTRODUCTION

Internet technology has a lot of benefits. It is also recognized as the fundamental human right. The Internet has also proven to be highly dynamic means for the communication. One of the biggest technology of the Internet is our social media technology. Social media comes in many forms including blogs, photo sharing platforms, forums, business networks, social gaming, chat apps, social networks,etc.

Terrorist organizations which are active on social media to promote their views and spreading propaganda are often leaderless. That makes tackling terrorist propaganda a difficult task. They used social media to recruit new members from all over the world. They also spread fear among the people of different countries by using Twitter and other social media platform.Among all social media platforms, Twitter is more public. When a tweet is posted on Twitter it can reach more people than other social media platforms. So, Twitter is the first choice of terrorist organizations to spread their views and propaganda. In the last two years, Twitter has suspended around million twitter account for spreading terrorism. Their approach of suspending an account is if someone reports any account, then they suspend the account if the account really promotes terrorism. It is a very time-consuming process. The motivation behind this project is to crawl tweets from the twitter in real-time and then analyze the tweets to determine the support of terrorism. This system will significantly improve the time to review the tweets supporting terrorism.

A. The main features :

1. Countering terrorism and protecting human rights as mutually reinforcing goals.
2. Countering violent extremism and radicalization that lead to terrorism.
3. The development of co-operative, community-oriented approaches to counter terrorism.
4. Community policing as part of a comprehensive, human rights-compliant strategy to prevent terrorism.

II. LITERATURE SURVEY

Rupali Patil al. [1] The procedure of sentiment analysis and its visualization is explained in detail concerning the topic Article 370. Sentiment analysis and opinion mining require detailed knowledge of how twitter and its python client Tweepy works to obtain the results. Python libraries like matplotlib and pandas are also used for simpler analysis and visualization of the tweets acquired. As a whole, from this paper it can be concluded that Pakistan is comparatively more concerned about the impact on its trade and has been somewhat more cynical when it comes to the sentiments of its Twitter users whereas India, on the other hand, is more concerned about increase in terrorism with a slightly positive attitude towards the revocation from the country's Twitter users. Therefore, the plots presented here accurately displays users' opinions during such a historical change in a Country's policy.

Babak Yadranjiaghdam [2]we have proposed a framework for realtime analysis of Twitter data. This framework is designed to collect, filter, and analyze streams of data and gives us an insight to what is popular during a specific time and condition. The framework consists of three main steps; data ingestion, stream processing, and data visualization. Data ingestion is performed by Kafka, a powerful message brokering system to import tweets, and to distribute it based on Topics that it defines, and to make it available over consumers' nodes to be processed by analytical tools. Apache Spark is used to access these consumers directly and analyze data by Spark Streaming. This allows not only general processing

tasks but more sophisticated and highlevel data analytics and machine learning algorithms..

M. Ashcroft et al. [3] made an attempt to detect jihadist messages from Twitter. They used sentiment analysis to detect if a message supports ISIS or not. They used some keyword to extract tweets from the Twitter feed. The advantage of this work is it uses three different features such as Time based features, Sentiment based features and Stylometric features to detect jihadist text. They got almost 90% accuracy using these features. One of the limitations of this study is they didn't use any real-time validation of their classification algorithm. Also, they didn't build any tools to detect jihadist text automatically.

Priyanka Harjule al.[4] Various techniques for both lexicon-based and machine learning based, have been applied in this project and the results are compared. It has been observed that for a totally new data/text machine learning based models trained over a related data are much more accurate than the classification based on standard dictionaries. This is because of the fact that the text that's being observed i.e the tweets are highly informal and do not use the standard grammar rules or the spelling and thus the data here is highly unstructured. The comparison results can be clearly observed among different machine learning algorithms also. As of now, among the algorithms used, RNN is observed to have the highest accuracy.

Walid M. et al. [5] studied to predict future support or opposition for ISIS from tweets. In this study, the authors used Twitter data to study the ISIS support of users. They used the bag of words model as feature vector which included individual terms, user mentions and hashtags. They used SVM with a linear kernel to train a classifier to predict the support or opposition of ISIS. They obtained about 87% accuracy using the SVM classifier model. One limitation of their approach is, they didn't consider real-time validation of their method.

S. Azrinaet al. [6] studied to detect terrorism from text using sentiment analysis. The advantage of their study is they did a comparative analysis on several techniques to detect terrorism from the text. They did a comparative analysis of Neural Network, Support Vector Machine, Sentiment Analysis with Naive Bayes and Lexicon based approach. Then they finally adopted and improved Naive Bayes method for their research. One of the limitations of this study is it uses user behavioral analysis to improve the accuracy of Naive Bayes algorithm as it shows medium level accuracy compared to other algorithms. But it is not always possible to analyse a user's behavior if he doesn't have that much of tweet history on Terrorism.

Lisa K. [7] studied to classify tweeps and tweets as being multipliers of jihadism. They used Machine Learning to build a classifier that can analyze a tweet to find multipliers of jihadism. They used AdaBoost classifier to train a model. They analyzed both Arabic and English tweets. They obtained about 84% accuracy for Arabic tweets and 98% accuracy for English tweets. They didn't test their model in a real-time environment.

Pooja W. et al. [8] studied to classify radical tweets in the categories such as Media, War terrorism, Extremism, Operations, Jihad, Country and Al-Qaeda. They built a dictionary to classify the tweets into different categories. They built the dictionary by looking at tweets containing hashtags like Al-Qaeda, Jihad, Terrorism, and Extremismand by collecting relevant words for their purpose. They built a process based on the presence of the word in the dictionary. Their process obtained about 90% accuracy. Their study

is proof that keywords can be used successfully in classifying tweets.

Lee S. et al. [9] compared four text mining methods: Latent Semantic Analysis (LSA), Probabilistic Latent Semantic Analysis (PLSA), Latent Dirichlet Allocation (LDA), and Correlated Topic Model (CTM) using topic model and spam filtering. They concluded that PLSA shows the highest performance and next to LDA, CTM, and LSA in order. One of the limitations of this study is they only considered statistical approach and didn't extend their study to syntactical and morphological approach.

A work on topic discovery based on text mining techniques was presented by Pons P. et al. [10]. They proposed a hierarchical clustering algorithm that combines partitioned and agglomerative approaches to produce topic hierarchies. They considered document place, time reference, and textual contents. This resulted in less time complexity while detecting a new topic. The accuracy of their proposed method is not satisfactory.

III. SYSTEM ARCHITECTURE AND DESIGN

The system architecture of Twitter Terrorism Detection Framework comprises five basic modules:

1) Twitter data crawler module
2) Storage module,
3) tweet classification module
4) Output module,
5) Training module.

A. Crawling Data from Twitter

We have crawled real-time tweets from the twitter by using twitter streaming API. To do this we have provided four keys that we have collected from the Twitter developer website.The 4 keys are:

1. Access Token
2. Access Secret Key
3. Consumer Token
4. ConsumerSecret Key.

These 4 key are needed to get access to Twitter API. Using these 4 key we can set up a twitter real listener. This listener will allow us to collect real-time tweets from Twitter. As we have used python to build our framework, we have used a python library that helped us getting access to the Twitter streaming API. We have used a python library 'Tweepy' API.Algorithm 1 illustrates the Twitter. While collecting real-time tweets we have made sure to handle all kinds of error that would break the connection the connection breaks, the crawler module will stop working. That's why we have checked for several exceptions such database exception, limit exceeded exception etc.

Algorithm 1: Crawl Real-Time Tweets

Input: Developer Access keys
Require: Real-Time tweets streaming from twitter

1. **Begin**
2. **Call** Twitter API
3. **Call** Twitter Streaming API
4. **Set**accessTkn = ""
5. **Set** accessTknSec = ""
6. **Set**consumerKey = ""
7. **Set**consumerSec = ""
8. tweets = STREAM-LISTENER(accessKeys)
9. **Create** a table named all_tweets having the field username, tweet, tweet_id, type
10. **if** tweets != null **then**
11. **if** tweets['retweet'] = False **then**
12. **Insert** tweets['text'] , tweets['username'] and tweets['tweet_id'] into the database
13. **End**

B. Pre-Processing Crawled Data

After data is crawled from the twitter the tweets are in raw form using algorithm 2. We can't use these tweets to classify or train. So, we have cleaned the tweet before using n classification or training. To handle the special component of a tweet, we have done the following pre-processing tasks.

1. URL is removed
2. Any user mention is removed
3. Hash (#) from the hashtag is removed
4. Contracted words are converted to their long form
5. Tokenized the tweet

Algorithm 2: Cleaning raw tweets.

Input: raw tweets
Require: clean the raw tweets

1. **Begin**
2. remove url from raw tweets
3. remove hash (#) symbol of hashtags from raw tweets
4. remove user mentions form raw tweets
5. remove retweet symbol RT from raw tweets
6. convert the raw tweets into lowercase form
7. search for contracted from in tweets
8. **if** contracted form found **then**
9. replace it with long form
10. search for stop words in tweets
11. **if** stop words found **then**
12. remove the stop words
13. tokenize the tweets
14. apply stemming on the tweets
15. **End**

C. *Building Model and Generating Output*

To predict the class of the tweet we needed a mathematical model which can classify the tweets based on their features. We have used two classification algorithm. These are SVM (Support Vector Machine) and Logistic Regression. Using our training dataset we built a model that can classify the tweets accurately. By using the model that we have built in the previous steps, we can classify a tweet. The classification result is 0 or 1 or 2. According to this result, we can show the type of tweets. Algorithm 3 is used to classifyreal-time tweets.

Algorithm 3: Classification of real-time tweets

Inputs: model file
Require: Classification of the tweets

1. **Begin**
2. classifier = load(model)
3. for each tweets in the all_tweets table in the database **do**
4. clean the tweets
5. type = classifier.predict(clean_tweets)
6. **if** type = 0 **then**
7. result = “Terrorism Supporting”
8. **elseif**type = 1 **then**
9. result = “Terrorism Non-Supporting”
10. **elseif** type = 2 **then**
11. result = “Random”
12. show the result
13. **End**

IV. CONCLUSION

In this Paper, We have discussed Twitter Terrorism Detection framework to detect tweets that support terrorism from real-time tweets stream. Our framework collects real time tweets by using twitter streaming API and analyses every tweet. It can categorize the tweet into three different classes and based on the category of the tweet, it is stored and shown in the different screen of our web application. We have also created a re-train module which will be used to retrain our model so that it can perform more accurately. Our framework has a user-friendly interface. The overall accuracy of our proposed system is 95% and 94% for SMV and Regression

respectively. In our study, the framework is limited to collect and analyze tweets that are written in English. So, further research can be done to extend this study to support other languages. The accuracy of the proposed system can be enhanced by analyzing shared images or videos on Twitter by users. Our framework will falsely detect sarcastic text as a terrorism-supporting tweet. So, further research can include detecting sarcastic tweets which actually doesn't support terrorism.

V. REFERENCES

[1]. Twitter Data Visualization and Sentiment Analysis of Article 370 ,2019.Author: Rupali Patil , Nishant Gada , Krisha Gala

[2]. Developing a Real-Time Data Analytics Framework for Twitter Streaming Data,2017.Author: Babak Yadranjiaghdam , Seyedfaraz Yasrobi , Nasseh Tabrizi

[3]. M. Ashcroft, A. Fisher, L. Kaati, E. Omer, and N. Prucha, “Detecting jihadist messages on twitter,” in Intelligence and Security Informatics Conference (EISIC) European, Sept 2015, pp. 161–164.

[4]. Text Classification on Twitter Data ,2020.Author:Priyanka Harjule , Astha Gurjar , Harshita Seth , Priya Thakur

[5]. M. Walid, D. Kareem and W. Ingmar, “ #FailedRevolutions: Using Twitter to Study the Antecedents of ISIS Support,” in Monday, 2015.

[6]. S. A. Azizan and I.A. Aziz, "Terrorism Detection Based on Sentiment Analysis Using Machine Learning," Engineering and Applied Sciences 2017.

[7]. K. Lisa, "Detecting multipliers of jihadism on twitter." International Conference on Data Mining Workshop (ICDMW) IEEE, 2015.

[8]. P.Wadhwaand M. P. S. Bhatia, “ Case study in Computing Achievements and Trends, Radical Messages onTwitter Using Security Associations 273. 2014.

[9]. P.P. Aurora, R.B. Llavori, and based on text mining techniques." &management, vol-43, no-3, pp.752

[10]. M. Trupthi, S. Pabboju and G. Narasimha, “Sentiment Analysis on Twitter Using Streaming API”, Computing Conference, 2017.

[11]. S. Lee, J. Baker, J. Song and comparison of four text mining methods”. In International Conference on 2010.

[12]. Tweepy, Streaming With Tweepy onlineTweepy.readthedocs.io. Available at: http://tweepy.readthedocs.io/en/v3.5.0/streaming_how_to.html 2017

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Survey on Real Time Road Lanes Detection of Autonomous Vehicles

Divya Sathe[1], Sayali Mhaske[1], Kunal Milkhe[1], Swapnil Nangre[1], Dr. Pankaj Agarkar[2], Prof. Pooja Shinde[3]

[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra, India

[2]HOD, Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra, India

[3]Assistant Professor, Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra, India

ABSTRACT

Autonomous road vehicles are increasingly becoming more important and there are several techniques and sensors that are being applied for vehicle control. Autonomous vehicles, Intelligent and Advanced Driving Assistant Systems are promising and reliable solutions to enhance road safety, traffic issues and passengers' comfort. An increasing safety and reducing road accidents, thereby saving lives are one of great interest in the context of Advanced Driver Assistance Systems. Apparently, among the complex and challenging tasks of future road vehicles is road lane detection or road boundaries detection. However, lane detection is a difficult problem because of the varying road conditions that one can encounter. Such applications require advanced computer vision algorithms that demand powerful computers with high speed processing capabilities. Keeping intelligent vehicles on the road until its destination, in some cases, remains a great challenge, particularly when driving at high speeds. The first principle task is robust navigation, which is often based on system vision to acquire RGB images of the road for more advanced processing. The second task is the vehicle's dynamic controller according to its position, speed and direction. In this paper we survey the approaches and the algorithmic techniques devised for the various modalities over the last 5 years. We present a generic break down of the problem into its functional building blocks and elaborate the wide range of proposed methods within this scheme.

Keywords: Advanced Driving Assistant Systems, lane detection, Autonomous vehicles.

I. INTRODUCTION

In real time lane detection system for autonomous vehicles, a lane detection and changing system is to warn and notify the vehicle driver when the vehicle is about to cross the lane and its dedicated path without the signal to turn. These terms are designed in such a way that it can reduce accidents, traffic and other circumstances where the driver is not paying attention or is distracted by phone call or other things. These mechanisms are totally beneficial for road management and traffic controlling at a greater part. In this paper, we have introduced a computer vision-based technique that can perfectly detect the lanes in

any suitable environment. Since most lanes on the road have clear lines whereas most of them are straight lines so that it is easy to detect the lanes and the lane detection technology for proper roads has reached a high milestone during these recent years. However, due to irregular surface and curved shape of the roads, and the unstructured roads are vulnerable to light, shadow, water and other factors, actors that result in poor detection performance. Therefore, the unstructured and irregular road lane detection technology is still in the research phase. The main deadlock in the implementation of such mechanisms is the prediction of problems which has two of the following factors: road lane assumption and obstacle assumption (i.e. vehicles and hurdle objects) detection. In this mechanism study we consider the first. Colour of road and texture, boundaries of roads, and lane markings are the main considerable criteria for human driving process. Semiautomated and fully automated vehicles are expected to share the same road and path with human drivers and therefore are most likely to rely continuously on the same assumed path that humans follow. While there can be different infrastructure assuming human drivers and vehicles like lane marks for humans and some kind of vehicle-infrastructure communication for vehicles it is unethical to expect the huge investments required to construct and maintain such infrastructures and road mechanisms, with the associated risk in mismatching and misguided marking .Road and lane presumption through the traditional ways remains therefore the most likely path for autonomous driven vehicles. Roads and lanes understanding includes detection the extent of the road, the number and position of lanes, merging, dividing and ending lanes and roads, in cities, villages and highways scenarios. Although there has been much progress in recent years, this type of understanding is beyond the reach of current detection systems. There are many detection modules used for road and lane understanding and detection, including monocular vision which is one video camera, stereo, LIDAR, vehicle dynamics information which is obtained from the car odometer or inertial measurement unit (IMU) with global positioning information obtained from global positioning system (GPS) and digital maps. Vision is the most primary research area in lane and road detection mechanism due to the fact that markings are made for human vision, while LIDAR and global positioning are important complements.

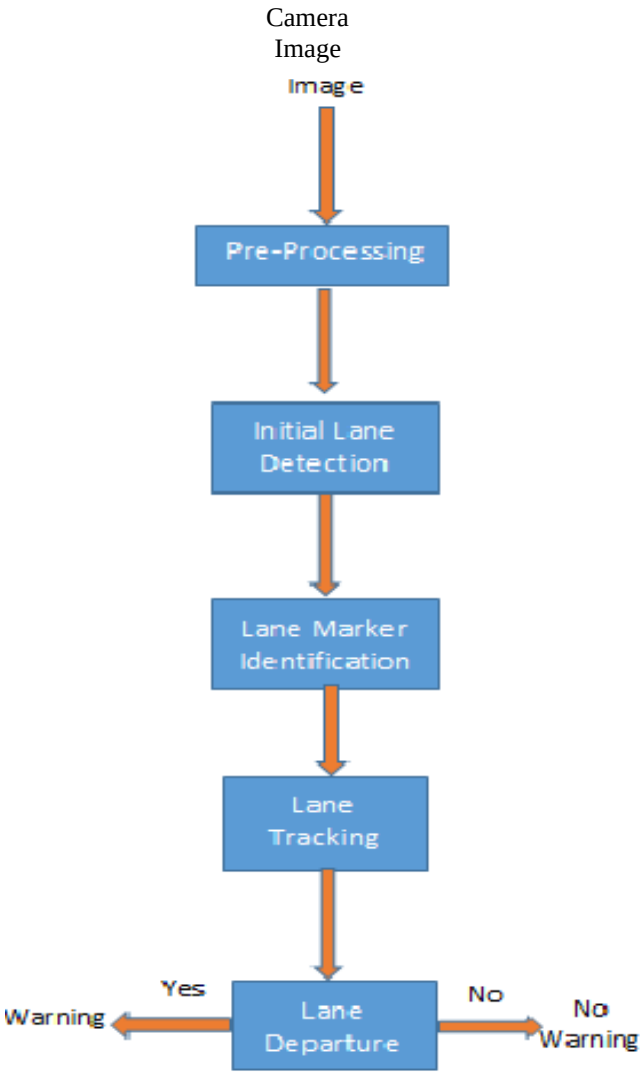


Figure.1. Block diagram of a simple Lane Departure Warning System (LDWS)

II. SURVEY ON LANE DETECTION ALGORITHMS

Optimal straight-line detection is a considerable step for several embedded vision applications, and now the largest research focus is based on the Hough Transform (HT). The straight line detection has been widely used in many industrial applications like image analysis, smart robots, intelligent vehicles, and pattern recognitions. Now a day's we observe that most researchers focused on increasing need for traffic safety systems to reduce the risk of accidents. There are large numbers of vision based systems for vehicle control, collision avoidance and lane departure warning, which have been developed during the last two decades. Recently, many Driver Assistance systems (DAS) are emerging to work in harmony with human drivers, e.g. Forward Collision Warning System (FCWS) and on-board Lane Departure Warning System (LDWS). Such systems can be used to help preventing driver's mistakes and reduce traffic accidents effectively.

Many researchers have worked on lane detection algorithm techniques. Various methods for lane detection algorithm are conferred as follow:

1) Y. Wang, E. K. Theo, and D. Shen introduced lane detection and tracking of Autonomus vehicles using b-snake. Using B-snake technique the lane detection and tracking of vehicles was done without using any cameras parameters. B-Spline can form any arbitrary shape by a set of control points. B-Snake based lane model is able to describe a wider range of lane structures. By using the knowledge of the perspective parallel lines the problems of detecting both sides of lane markings have been merged here as the problem of detecting the mid-line of the lane. This method is robust against noise, shadows, and illumination variations in the captured road images. It is applicable to the dash and the solid paint line roads also to the marked and the unmarked roads.

2) M. Aly proposed a Real time detection of lane markings in urban streets. It is a real time, efficient and robust algorithm in urban streets for detecting lanes. In this algorithm firstly a top view of the road is produced and then it is filtered using various selective which is associated with Gaussian filters. In this algorithm there is a work on the red channel, which gives better images for white and yellow lanes than converting it to gray scale.

3) Kim developed a lane detection and tracking algorithm which is capable of handling various challenging scenarios such as faded lane markers,lightly marked lines, lane curvatures and splitting lanes. Firstly , a gradient detector and an intensity bump detector is used to removethe non-lane markers. Artificial Neural Networks (ANN) is applied on remaining samples for lane detection. The detected lane markers pixels are grouped together using cubic splines. Hypotheses are produced from random set of line segments. RANSAC algorithm helps in validating the hypotheses. In this paper Particle filtering was also used for lane tracking.

4) Yim and Oh developed a three feature based lane detection algorithm. There were various features used in this paper like starting position, orientation and intensity value. Firstly, a Sobel operator is applied to get the edge information. The lane boundary is represented as a vector comprising of the three features. The current lane vector is calculated based on the input image and the previous lane model vector. Two windows, one for each, is used for left and right boundaries. Assuming N pixel in each horizontal line, N lane vector candidates are generated. The best candidate is selected based on the minimum distance from previous lane vector using a weighted distance

5) metric. For equalization each feature is assigned a different weight. Then a lane inference system is used to predict the new lane vector. If the road width changes abruptly, the current vector calculated is discarded and the previous one is taken as current vector.

6) Borkar et al. developed a method based on the parallel nature of lane markers. Firstly the Inverse Perspective Mapping is performed. Then the IPM image is converted to gray scale image. Then the image is filtered using Normalized Cross Correlation. Now find out a collection of straight lines using Polar Randomized Hough Transform .To determine if two lines are parallel lines peaks with identical θ value can be paired. Usually the lane markers are parallel but due to some imperfections in lens, captured image variation in lane marker placements etc. So the constraint of identical θ value needs to be loosened. This can be achieved by applying a tolerance window. The video is tested in real time videos and obtained good results. Common difficulties face in lane detection such as presence of shadows neighbouring vehicles and surface irregularities are greatly reduced in this approach. There is difficulty in detecting worn out lane markers.

III. COMPARISON OF LANE DETECTION ALGORITHMS

Sr.No	Used Method	Advantages	Disadvantages
1.	Lane Detection And Tracking Using B-Snake.	This algorithm is proposed without using any cameras parameters	Processing time depend on the number of edge pixels and pre-specified road model
2.	Real Time Detection of Lane Markers in Urban Streets	It is a real time, efficient and robust algorithm in urban streets for detecting lanes	It does not give well accurate results
3.	Robust Lane Detection And Tracking In Challenging Scenarios	Deals with a lane curvature, lane changes, worn lane markings and emerging, merging, ending, and splitting lanes.	Faulty detection in curvy roads and wrong detection due to shadows
4.	Lanes Detection Based on	algorithm can be implemented	It cannot detect changes in line

	Unsupervised And Adaptive Classifier and pipelining structure	in smart phones	
5.	Hough Transform Algorithm for Lane Detection	It is robust against noise, high speed algorithm	High computational complexity

Table.1 Comparison of different techniques

IV. DRAWBACKS OF VARIOUS PAPERS

1. Gurveen Kaur proposed Lane detection Techniques: Here various techniques such as Canny Edge detector, Hough Transform are used. It is not suitable for various environmental conditions like rainy day, foggy day etc.

2. Farid Bounini, Denis Gingras proposed Autonomous Vehicle and Real Time Road Lanes Detection and Tracking. Here Least Square method, Kalman Filter techniques were used. The disadvantage of this system is it can work for a vehicle with a maximum speed of 70 km/h.

3. M. Aly proposed Real time detection of lane markers in urban streets.RANSAC fitting technique is used in this proposed system. The drawback of this algorithm is it does not give accurate results for the lane detection.

4. Zu Andras F. Cela, Luis M. Bergasa, Franklin L. Sanchez and Marco Aa introduced Lanes Detection Based on Unsupervised and Adaptive Classifier. Hough transform algorithm and Kalman filters are used. The drawback is that it cannot detect the changes in line.

5. Zu Whan Kim proposed Robust Lane Detection and Tracking in Challenging Scenarios. Lane-detection and-tracking algorithm is used in this system. The drawback of this system is there is faulty detection in curvy roads and wrong detection due to shadows.

6. Hyo-Kyun Jeong, Yong-Jin Jeong, and Yi-Fan Lin proposed a design of Hough transform hardware accelerator for Lane detection. Hough transform algorithm is used here. The disadvantage of Hough transform is it has high computational complexity.

V. CONCLUSION

In this review, a detailed analysis of various lane detection and tracking algorithms is discussed. The different methodologies investigated by different authors for lane detection and tracking during the last decade are presented in the paper. Lane detection is important because it is an integral part of autonomous vehicle control system. A survey of existing methods for detection and marking of lanes is provided in this work. The previous methods proposed for detection and marking have several shortcomings. A lane departure warning system and lane marking detection using Hough Transform has been proposed. By incorporating a departure warning system, the functionalities of the lane marking system can be enhanced therefore, further improvements can be done to enhance the results. In the near future, one can modify the existing Hough Transformation so that it can measure both the curved and straight roads. Hough transform can be implemented on FPGA board. FPGA implementation consumes less power also it is very compact and fast.

VI. REFERENCES

[1]. Gurveen Kaur,"Lane Detection Techniques: A Review International Journal Of Computer Applications (0975-8887) Volume 112-No 10, February 2015

[2]. M. Aly, "Real time detection of lane markers in urban streets," IEEE Intelligent Vehicles Symposium, pp. 7-12, June 2008.

[3]. Zu Andras F. Cela, Luis M. Bergasa, Franklin L. Sánchez and Marco A. Herrera, "Lanes Detection Based on Unsupervised and Adaptive Classifier", 978-0-7695- 5042-8/13 $26.00 © 2013 IEEE.

[4]. Zu Whan Kim, Member, IEEE, Robust Lane Detection and Tracking in Challenging Scenarios", IEEE transactions on intelligent transportation systems, Vol. 9, No. 1, March 2008.

[5]. Hyo-Kyun Jeong, Yong-Jin Jeong, IEEE, and Yi-Fan Lin, "Design of Hough transform hardware accelerator for Lane detection", 978-1 4799-2827-9/13/$31.00 ©2013 IEEE.

[6]. A.Borkar, M. Hayes, M. Smith, and S. Pankanti, "A layered approach to robust lane detection at night," in Proc. IEEE Workshop Comput. Intel. Vehicles Veh. Syst., 2009, pp. 51-57.

[7]. A.Assidiq, O.khalifa, Islam and S.khan, "Real time lane detection for autonomous vehicles" in Computer and Communication Engineering, 2008. ICCCE 2008. International Conference on. IEEE, 2008, pp. 82-88.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Detection of Depression or Sentiment Analysis

Chaitanya Suryawanshi[1], Taufik Tamboli[1], Saurav Tayade[1], Prashant Yeole[1], Prof. Niyamat Ujloomwale[1]
[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

Depression is ranked as the largest contributor to global disability and is also a major reason for suicide. Still, many individuals suffering from forms of depression are not treated for various reasons. Previous studies have shown that depression also has an effect on language usage and that many depressed individuals use social media platforms or the internet in general to get information or discuss their problems. In particular, a convolutional neural network based on different word embeddings is evaluated and compared to a classification based on user-level linguistic metadata. An ensemble of both approaches is shown to achieve state-of-the-art results in a current early detection task. Furthermore, the currently popular ERDE score as metric for early detection systems is examined in detail and its drawbacks in the context of shared tasks are illustrated. A slightly modified metric is proposed and compared to the original score. Finally, a new word embedding was trained on a large corpus of the same domain as the described task and is evaluated as well. Social networks have been developed as a great point for its users to communicate with their interested friends and share their opinions, photos, and videos reflecting their moods, feelings and sentiments. This creates an opportunity to analyze social network data for user's feelings and sentiments to investigate their moods and attitudes when they are communicating via these online tools.

Keywords: Social network, Emotions, Depression, Sentiment analysis.

I. INTRODUCTION

According to World Health Organization (WHO) , more than 300 million people worldwide are suffering from depression, which equals about 4.4% of the global population. While forms of depression are more common among females (5.1%) than males (3.6%) and prevalence differs between regions of the world, it occurs in any age group and is not limited to any specific life situation. Depression is therefore often described to be accompanied by paradoxes, caused by a contrast between the self-image of a depressed person and the actual facts . Latest results from the 2016 National Survey on Drug Use and Health in the United States report that, during the year 2016, 12.8% of adolescents between 12 and 17 years old and 6.7% of adults had suffered a major depressive episode (MDE). Precisely defining depression is not an easy task, not only because several sub-types have been described and changed in the past , but also because the term "being depressed" has become frequently used in everyday language. In general, depression can be described to lead to an altered mood and may also be accompanied The proliferations of internet and communication technologies, especially the online social networks

have rejuvenated how people interact and communicate withbeach other electronically.

The applications such as Facebook, Twitter, Instagram and alike not only host the written and multimedia contents but also offer their users to express their feelings, emotions and sentiments about a topic, subject or an issue online. On one hand, this is great for users of social networking site to openly and freely contribute and respond to any topic online; on the other hand, it creates opportunities for people working in the health sector to get insight of what might be happening at mental state of someone who reacted to a topic in a specific manner. In order to provide such insight, machine learning techniques could potentially offer some unique features that can assist in examining the unique patterns hidden in online communication and process them to reveal the mental state (such as 'happiness', 'sadness' ,'anger', 'anxiety', depression) among social networks' users. Moreover, there is growing body of literature addressing the role of social networks on the structure of social relationships such as breakup relationship, mental illness ('depression', 'anxiety', 'bipolar' etc.), smoking and drinking relapse, sexual harassment and for suicide ideation The symptoms experienced by depressed individuals can severely impact their ability to cope with any situation in daily life and therefore differ drastically from normal mood variations that anyone experiences. At the worst, depression can lead to suicide. WHO estimates that, in the year 2015, 788,000 people have died by suicide and that it was the second most common cause of death for people between 15 and 29 years old worldwide In Europe, self-harm was even reported as the most common cause of death in the age group between 15 and 29 and the second most common between 30 and 49, again in results obtained by WHO in 2015. In this study, we examine various linguistic cues which help to detect emotion cause events: the position of cause event and experiencer relative to the emotion keyword: emotional process like positive emotion (e.g.'happy', 'love', 'nice'), negative emotion (e.g. 'worthless ','loser', 'hurt', 'ugly', 'nasty'), sadness (e.g. 'worry', 'crying' ,'grief', 'sad'), anger (e.g. 'stop', 'shit', 'hate', 'kill', 'annoyed')and anxiety (e.g. 'worried', 'fearful'). A temporal process like present focus (e.g. 'today', 'is', 'now'), past focus (e.g. 'ago', 'did', 'talked') and future focus (e.g. 'shall', 'may', 'will' ,'soon'). Linguistic words like articles (e.g. 'a', 'an', 'the'),prepositions (e.g. 'for', 'in', 'of', 'to', 'with', 'above'), auxiliary verbs (e.g. 'do', 'have', 'am', 'will'), conjunctions (e.g. 'and' ,'but', 'whereas'), personal pronoun (e.g. 'I', 'them', 'her' ,'him'), impersonal pronouns (e.g. 'it', 'it's', 'those'), verbs (e.g. 'go', 'good') and negation (e.g. 'deny', 'dishonest', 'no ','not', 'never').

II. LITERATURE REVIEW

Previous studies have already shown that depression also has an effect on the language used by affected individuals. For example, a more frequent use of first person singular pronouns in spoken language was first observed in 1981. An examination of essays written by depressed, formerly-depressed, and non-depressed college students at University of Texas confirmed an elevated use of the word "I" in particular and also found more negative emotion words in the depressed group. Similarly, a Russian speech study found a more frequent use of all pronouns and verbs in past tense among depression patients. A recent study based on English forum posts observed an elevated use of absolutist words (e.g. absolutely, completely, every, nothing1) within forums related to depression, anxiety, and suicidal ideation than within completely unrelated forums as well as ones about asthma, diabetes, or cancer .The knowledge that language can be an indicator of an individual's psychological state has, for example, lead to the development of the Linguistic Inquiry and Word Count (LIWC) software.

By utilizing a comprehensive dictionary, it allows researchers to evaluate written texts in several categories based on word counts. A more detailed description of LIWC. With a similar purpose, Differential Language Analysis Toolkit (DLATK) an open-source Python library, was created for text analysis with a psychological, health, or social focus. Driven by the growing availability of data, for example through social media, and the technological advances that allow researchers to work with this data, ethical considerations are becoming more and more important in the field of Natural Language Processing (NLP). Based on these developments, NLP has changed from being mostly focussed on improving linguistic analysis towards actually having an impact on individuals based on their writings. Still, a proper discussion about ethics in NLP has only been started in 2016 by Hovy and Spruit . Although Institutional Review Boards (IRBs) have been well-established to enforce ethical guidelines on experiments that directly involve human subjects, the authors note that NLP and data sciences in general have not constructed such guidelines. They further argue that language "is a proxy for human behavior, and a strong signal of individual characteristics" and that, in addition, "the texts we use in NLP carry latent information about the author and situation" . On top of this direct connection to the individual, they also describe the social impact of NLP research . A demographic bias in the selection of training texts can lead to the exclusion of specific groups, overgeneralization based on false positives can have serious consequences depending on the task, and research results can potentially cause or confirm biases and ultimately discrimination by topic overexposure. Even if all these factors are considered, they conclude that dual-use problems can exist for any research if results are used in a different way than originally intended.

The work described in this paper belongs to the area of Natural Language Processing (NLP) and text classification in particular. The origins of text classification tasks can be found in early research to automatically categorize documents based on statistical analysis of specific clue words in 1961 . Later, similar research goals lead to rule based text classification systems like CONSTRUE in 1990 and finally the field began to shift more and more to machine learning algorithms around the year 2000 In addition to text categorization, machine learning was also a driving force in other text-based tasks like sentiment analysis, which is focussed on extracting opinions and sentiment from text documents . It was first used in combination with machine learning to find positive or negative opinions in movie reviews and was then extended to other review domains , as well as completely different areas like social media monitoring and general analysis of consumer attitudes. More recently, deep learning has been utilized for text classification in addition to its more common, usages in image classification. State-of-the-art results in several text-based tasks could, for example, be achieved by transfer learning methods like Universal Language Model Fine-tuning (ULMFit) and the Google research project Bidirectional Encoder Representations from Transformers(BERT) for the training of language representations, which includes ULM Fit and several other methods. Based on these Especially the availability of social media messages enabled researchers to extract population-based health information that made it possible to track disease symptom.

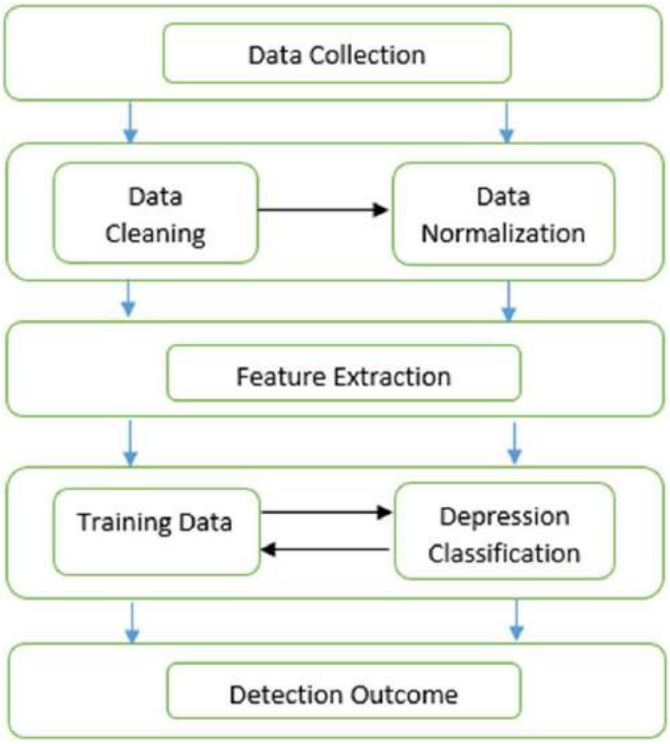


Fig. 1 A methodological overview of Facebook data analysis for depression analysis

There is growing body of literature that analyses the properties of depression . Choudhury et al.
argue that depression constitutes a genuine test in individual and general wellbeing. Considerable number of individuals experiences the ill-effects of despondency and just a division gets sufficient treatment every year. They also investigated the possibility to utilize online networking to identify and analyze any sign of significant depression issue in people. Through their web-based social networking postings, they quantified behavioral credits identifying with social engagement, feeling, dialect and semantic styles, sense of the self-system, and notices of antidepressant medications.
Choudhury et al. considered online networking
as a promising instrument for public health, concentrating on the utilization of Twitter presents on fabricating predictive models about the forthcoming impact of childbirth on the conduct and disposition of new mothers. Utilizing Twitter posts, they measured postpartum changes in 376 mothers along measurements of social engagement, feeling, informal community, and phonetic style. O'Dea et al. examined that Twitter is progressively researched as methods for recognizing psychological well-being status, including depression and suicidality in the population. Their investigation revealed that it is conceivable to recognize the level of worry among suicide- related tweets, utilizing both human coders and a programmed machine classifier.
Zhang et al. have shown that if individuals with a
high danger of suicide can be recognized through online Islam *et al. Health Inf Sci Syst (2018) 6:8* Page 3 of 12 networking like microblog, it is conceivable to actualize a dynamic intervention system to save their lives. Many researchers have demonstrated that utilizing user-created content (UGC) accurately may help decide individuals' psychological wellness levels. For instance, Aldar wish and Ahmad examined that the utilization of Social Network Sites (SNS) is expanding these days, particularly by the more youthful eras. Because the accessibility of SNS enables clients to express their interests, sentiments and offer day by day schedule .
Nguyen et al. utilized machine learning and statistical strategies to separate online messages amongst depression and control groups utilizing temperament,
psycholinguistic procedures and substance subjects
removed from the posts created by individuals from these groups. Park et al. investigated states of mind and practices toward online web-based social networking in view of whether one is discouraged or not. They directed semi-organized up close and personal meetings with 14 dynamic Twitter users, half of whom were discouraged and the other half non-discouraged. Other than they examined a few plan implications for future social networks that could better suit users with depression and give bits of knowledge towards helping discouraged users address their issues through online web-based social

networking. Bachrach et al. studied how user's activity on Facebook identifies with their identity, as measured by the standard Five Factor Model. They analyse relationships between user's identity and the properties of their Facebook profiles. For instance, the size and thickness of their friendship network, number of transferred photographs, and number of occasions went to, number of gathering enrolment's, and number of times the user has been tagged in photographs.

Ortigosa et al. have exhibited a new strategy for sentiment examine in Facebook that suggests that starting from messages composed by users,

as to extract data about the users' assessment extremity (positive, unbiased or negative), as transmitted in the messages they write; and to show the users' standard conclusion extremity and to distinguish huge passionate changes. In the context of Facebook mining, Holleran found initial evidence that depression is a major contributor to the overall global burden of diseases. In other related work, Wang et al. and Shen et al. examined various depression-related features, and built amultimodal depressive model to detect the depressed users. Although, some of the above reported work has discussed emotional process, temporal process, linguistic style to detect depression, the following shortcomings are observed in the existing literature:

There are few individual studies that have applied SVM, KNN, Decision Tree and Ensemble separately. There are no well-known studies that have combined all these techniques together at same dataset to investigate the variations in technique-based findings. There is no significant study that has applied the abovementioned machine learning techniques on Facebook data for depression detection. To address the above-listed shortcomings, we make an attempt to detect depression from Facebook comments with the present work; expand the scope of social media-based depression measures, describing the different features of Facebook user comments. applied machine learning approaches that can use those measures for the detection of individuals who are suffering with depression.

Year	authors	Data
1981	University of Texas	more frequent use of first person singular pronouns in spoken language
2017	Almeida, H., Briand, A., Meurs, M.J	Detecting early risk of depression from social media user-generated content. In: Proceedings Conference and Labs of the Evaluation Forum CLEF
2018	Cacheda, F., Fernandez, D., Novoa, F., Carneiro, V.:	Artificial intelligence and social networks for early detection of depression.
2017	Trotzek, M., Koitka, S., Friedrich, C.M	Linguistic metadata augmented classiers at the clef 2017 task for early detection of depression. In: Proceedings Conference and Labs of the Evaluation Forum CLEF
2014	Prieto, V.M., Matos, S., Alvarez, M., Cacheda, F., Oliveira, J.L.:	Twitter: a good place to detect health conditions.
2017	Aldarwish MM, Ahmad HF	Predicting depression levels using social media posts. In: 2017 IEEE 13th international Symposium on Autonomous decentralized system

Table 1. Summary of literature review

III. SYSTEM ARCHITECTURE

There are a number of ways to analyze the information, but the reality is that mental health, specifically depression, is a subjective and complex topic. While it may be possible to quantify the degree to which one might be depressed based on a Tweet, the only real question that matters for this project is, is an individual exhibiting linguistic markers indicative of depression? Knowing the question and the subjective nature of mental illness, a binary classification model made the most sense for this project. While a logistic regression model made sense as a benchmark model, a Long Short Term Memory network (LSTM) model wound up being the most robust for the project at hand. A recurrent neural network allows information to be passed from one

step of a network to another, and are ideal for sequences, lists, and other language processing problems. A LSTM is capable of learning long-term dependancies and work incredibly well on a large variety of problems. The LSTM + CNN model takes in an input and then outputs a single number representing the probability that the tweet indicates depression. The model takes in each input sentence, replaces it with its embeddings, and then runs the new embedding vector through a convolutional layer. The convolutional layer passes the structure that it learns from the sequential data into a LSTM layer. The output of the LSTM layer is then fed into a Dense model for prediction.

Once the model was designed and built, the issue then became refining the model to achieve the best results.

Fig. 2 A methodological overview of tweets analysis for depression analysis

The system will send the tweets to be analyzed and stored the results in the database. The tweets will be analyzed in all three models. The system will return the predicted sentiments which are Positive, Negative or Neutral. When the system returns two Positive results and one Negative or Neutral result, the system will take the Positive predicted sentiment as for the Overall Predicted Sentiment, same as for two Negative results and two Neutral results. The tweets are then analyzed using three different techniques which are Naïve Bayes Classifier technique, NLP techniques and Deep Learning technique

After the sentiment of each user tweets is calculated, the depression percentage is then calculated based from the total positive and total negative tweets. If the users have a high percentage of positive tweets, it will classify the users as an optimistic person that implies the user is no depression related. Meanwhile, users that have a high percentage of negative tweets, it will classify the users as an optimistic person that can implies the users might be depression related

IV. CONCLUSION

In this paper exhibited the capability of using or measuring and detecting major depression among its users. To give a clear understanding of our work, numbers of research challenges were stated at the start of this paper. The analytics performed on the selected dataset, provide some insight on the research challenges : What depression is and what are the common factors contributing toward depression. While we feel moody, sad or low from time to time, few people encounter these emotions seriously, for drawn out stretches of time (weeks, months or even years) and in some cases with no apparent reason. Despondency is something other than a low state of mind—it's a genuine condition that influences someone's physical and emotional feelings. Depression can influence any of us anytime. However, some phases or events make us more vulnerable to depression. Physical and emotional changes associated with growing-up, losing a loved one, beginning a family, retirement may trigger some emotional influx that could lead toward depression

for few people. What are the factors to look for depression detection in social networking comments? It is important to remember that depressive emotions have several signs and symptoms spread across various categories as reported in Based on signs and symbols divided dataset into 5 emotional variables (positive, negative, sad, anger, anxiety), 3 temporal categories (present focus, past focus and future focus), 9 standard linguistic dimensions (e.g., articles, prepositions, auxiliary verb, adverbs, conjunctions, pronoun, verbs and negations) How to extract these factors from social sites comments ? To extract the above-mentioned factors, we applied Linguistic Inquiry and Word Count (LIWC) on our dataset. The LIWC2015 Dictionary is the heart of the text analysis strategy. It processes our comments on a 'line by line' basis within and across columns of spreadsheet and accesses a single text within a spreadsheet and analyse each line sequentially and reads one target word at a time. What is the relationship between these factors and attitudes toward depression? The relationship between the above-mentioned issues with the attitudes towards depression are varies from person to person. When are the most influential time to communicate within depressive Indicative Facebook user? In this study, got 54.77% depressive indicative Facebook users communicate with their friends from midnight to midday and 45.22% from midday to midnight.

V. REFERENCES

[1]. Depression and Other Common Mental Disorders: Global Health Estimates. World Health Organization, 2017.

[2]. A. T. Beck and B. A. Alford, Depression: Causes and Treatment. Second Edition. University of Pennsylvania Press, 2009.

[3]. Key Substance Use and Mental Health Indicators in the United States: Results from the 2016 National Survey on Drug Use and Health. Rockville, MD: Center for Behavioral Health Statistics and Quality: Substance Abuse and Mental Health Services Administration, 2017. OnlineAvailable: https://www.samhsa.gov/data/

[4]. E. S. Paykel, "Basic concepts of depression," Dialogues in Clinicaln Neuroscience, vol. 10, no. 3, pp. 279–289, 2008.

[5]. Global Health Estimates 2015: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2015. World Health Organization, 2016.

[6]. J. Alonso, M. Codony, V. Kovess, M. C. Ange rmeyer, S. J. Katz, J. M. Haro, G. De Girolamo, R. De Graaf, K. Demyt tenaere, G. Vilagut et al., "Population level of unmet need for mental healthcare ineurope," The British Journal of Psychiatry, vol. 190, no. 4, pp. 299

[7]. P. S. Wang, M. Angermeyer, G. Borges, R. Bruffaerts, W. T. Chiu, G. De Girolamo, J. Fayyad, O. Gureje, J. M. Haro, Y. Huang et al., "Delay and failure in treatment seeking after first onset of mental disorders in the world health organization's world mental health survey initiative," World Psychiatry, vol. 6, no. 3, p. 177, 2007.

[8]. A. Rahman, S. U. Hamdani, N. R. Awan, R. A. Bryant, K. S. Dawson, M. F. Khan, M. M.-U.-H. Azeemi, P. Akhtar, H. Nazir, A. Chiumento et al., "Effect of a multicomponent behavioral intervention in adults impaired by psychological distress in a conflictaffected: A randomized clinical trial," JAMA, vol.316, no. 24, pp. 2609–2617, 2016.

[9]. G. Schomerus, H. Matschinger, and M. C. Angermeyer, "Them stigma of psychiatric treatment and help-seeking intentions for depression," European Archives of Psychiatry and Clinical Neuroscience ,vol. 259, no. 5, pp. 298–306, 2009.

[10]. R. Whitley and R. D. Campbell, "Stigma, agency and recovery amongst people with severe mental illness," Social Science &Medicine, vol. 107, pp. 1 – 8, 2014.

[11]. K. Gowen, M. Deschaine, D. Gruttadara, and D. Markey, "Young adults with mental health conditions and social networking websites:Seeking tools to build community." Psychiatric Rehabilitation Journal, vol. 35, no. 3, pp. 245–250, 2012.

[12]. Haberler G. Prosperity and depression: a theoretical analysis of cyclical movements. London: Routledge; 2017.

[13]. Guntuku SC, et al. Detecting depression and mental illness on social media: an integrative review. Curr Opin Behav Sci. 2017;18:43–9.

[14]. De Choudhury M, et al. Predicting depression via social Media. In: ICWSM, vol. 13. 2013. p. 1–10.

[15]. De Choudhury M, Counts S, Horvitz E. Predicting postpartum changes in emotion and behavior via social media. In: Proceedings of the SIGCHI conference on human factors in computing systems. New York: ACM; 2013.

[16]. O'Dea B, et al. Detecting suicidality on Twitter. Internet Interv. 2015;2(2):183–8.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Artificial Intelligence based COVID-19 classification by using Deep Learning and Convolutional Neural Network

Omkar Gaikwad[1], Divyanshu Tripathi[1], Madhuri Dange[1], Prof. Pallavi Shimpi[2]
[1,2]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

COVID-19 irruption has place the total world in associate unexampled troublesome state of affairs delivery life round the world to a daunting halt and claiming thousands of lives. because of COVID-19 unfold in 212 countries and territories and increasing numbers of infected cases and death tolls mounting to five,212,172 and 334,915, it remains a true threat to the general public health system. This paper renders a response to combat the virus through Artificial Intelligence(AI) primarily based respiratory organ illness Classification by victimization organic process Deep Learning commonplace. The design of the projected model initial goes through a pre-processing of the input image is followed by a rise in information. Then the model begins a step to extract the characteristics followed by the training step. Finally, the model begins a classification and prediction method with a totally connected network fashioned of many classifiers. The model explains associate integrated bioinformatics approach during which completely different} aspects of information of knowledge taken from different data sources area unit place along to make the easy platforms for physicians and researchers. the most advantage of those AI-based platforms is to accelerate the method of identification and treatment of the COVID-19 illness. the foremost recent free publications and medical reports were investigated to decide on inputs and targets of the network that would facilitate reaching a reliable Artificial Neural Network-based tool for challenges related to COVID-19.

Keywords: Artificial Intelligence, Covid-19, Deep Learning, Convolution Neural Network, Image Processing

I. INTRODUCTION

The novel Coronavirus selected SARS-CoV-2 appeared in December 2019 to initiate an outbreak of health problem called COVID-19 that established itself as a tough illness which will emerge in numerous forms and levels of severity starting from delicate to severe with the chance of organ failure and death. From mild, selflimiting tract ill health to severe progressive respiratory illness, multi-organ failure, and death. the planet Health Organization declared it as an outbreak on Gregorian calendar month thirty, 2020. With the progress of the pandemic and also the rising range of the confirmed cases and patients WHO expertise severe metabolic process failure and vas complications, there area unit solid reasons to be enormously involved concerning the results of this virus infection. decisive acceptable

approaches to achieve solutions for the COVID-19 connected issues has received an excellent deal of attention. Healthcare systems round the world are trying to expand testing facilities for COVID-19. additional and additional testing can cause the identification and isolation of infected persons, thereby reducing the unfold among the community. however availableness doesn't guarantee responsibility. the key concern for the governments at this stage is that the false-negative take a look at results area unit negative for the infected individual. Such people could unwittingly transmit the virus to others. False take a look at outcomes so damage the efforts to curb the unfold of the virus. thence the sensitivity of those tests is unknown. In such a state of affairs, Deep Learning techniques area unit artificial neural networks during which every layer has multiple neurons that operate equally to the neurons of the flesh. Convolution neural networks (CNNs) area unit one in all the deep learning techniques that have proved to achieve success and effective within the field of medical imaging classification. There are many studies that have used CNN to diagnose respiratory illness and different diseases supported radiography. CNN primarily based design has been projected to spot completely different respiratory organ diseases. thence CNN becomes a natural candidate for identification recommendation of

COVID-19 patients. By coaching convolutional neural networks (CNN) victimization these characteristics extracted from X-ray pictures, we tend to may accurately predict COVID19. The results area unit encouraging and demonstrate the effectiveness of deep learning, and additional specifically, transfer learning with CNN to the automated detection of abnormal X-ray pictures from little information sets, associated with the Covid-19 illness.

II. RELATED WORK

In[1],Recent study proposed several COVID-19 detection paradigms. In [2], Li et al. proposed a methodology to recognize the infection rate using the coronal and axial view of lung CT scans. The proposed work achieved a specificity of 100%, AUC of 0.918, and sensitivity of 82.6%. Another study evaluated COVID19 disease using visual inspection. They claimed that visual inspection help to identify the infection. In [3], Yang et al. proposed a scheme to evaluate the lung CT scans and implemented visual inspection-based detection. Their scheme could achieve a Specificity of 94%., an AUC of 0.892, and a Sensitivity of 83.3%. In [4], Wang et al. investigated 90 patients' lung CT scans. Their investigation managed to detect the severity based on the time since the patient got infected. Also, a diagnostic methodology was proposed based on the CT scan's image features. They concluded that the combination of both images features evaluation and clinical findings could early detect the presence of COVID-19. In [5], Bai et al. investigated the patient's information and considered the CT scans and RT-PCR for the examination. They achieved a specificity of 100% and a sensitivity of 93%. In a similar study, authors clinically evaluated patients with both CT scans and real-time RT-PCR with an early detection accuracy of 90%.

In [6], an in-depth study of various techniques used for the classification of images is performed. The existing DCNN models were used for the prediction of COVID19 using CT and CXR images. The analysis is stated in terms of accuracy for various prediction models. A comprehensive study is performed in the automation of DCNN architecture for searching and classification of images.

Zhou et al. in 2020 [7] suggested a deep learning model for distinguishing influenza pneumonia taken from CT images and novel coronavirus pneumonia.

CT images are better than CXR images as it shows pulmonary infection clearly which is much costlier. . Li et al. [8] identified COVID-19 using Artificial Intelligence (AI), thus dataset comprising of affected COVID-19 images, various pneumonia, and diagnosed patients with pneumonia. The images are gathered from Chinese hospitals containing 2969 images of the training set, viral pneumonia 1396, more than 400 images of COVID-19 patients, and 1173 non-pneumonia. K. He, X.
Zhang et al. in 2016.

In [9], Jakimovski and Davcev proposed the Double convolutional deep neural network (DCDNN) for lung cancer stage prediction. In the training of both the CDNN and the regular CDNNs, they used Computed Tomography (CT) scans. These topologies have been experimented with against images of pulmonary cancer to assess the stage of TX cancer in that topologies can predict lung cancer. The first phase involved the preclassification of CT images from the first dataset to concentrate on CDNN research. Next, they create the double deep neural network Convolution with max pool to carry out a more comprehensive search. They eventually utilized Computed Tomography scans of various stages TX cancer of lung cancer to establish the TX stage where CDNN detects lung cancer potential. The findings are reviewed with the medical doctors of the oncology department and are deemed sufficient to detect cancer in T3. This technique is used by doctors for detection and treatment. Nasser and Abu-Naser introduced the Artificial Neural Network (ANN) for detecting lung cancer. A Network for the Prediction of the Absence or Presence of Lung Cancer in the Human Body was developed. Symptoms like yellow fingers, anxiety, chronic illness, weariness, allergy, wheezing, roar, breath shorter, swallowing difficulty, and chest pain was used to diagnose lung cancer. Symptoms like yellow fingers. They were used as input variables for our ANN and other details about the human. Trained and validated data collection entitled' survey cancer of the lung.' The system evaluation showed that the ANN model would detect 96.67 percent accuracy in the absence or presence of lung cancer. They have carried out some planning and analysis to make the data more predictive.

III. METHODOLOGY

In this section, we first describe the dataset used in the study, followed by the proposed CNN. The definition of the input data and desired outputs before the actual methods provides a better definition of the problem and a better understanding of the methods.

A. Datasets

We set up a database composed of three classes of chest radiographic images are collected from git hub repositories [10]. The first class is made up of images of patients declared positive for COVID-19 that we collected from the database published by Cohen. This database contained 230 images is open to various researchers to add new images or to use the already existing images. The second class consists of 100 images of patients declared normal without any pneumonia. The third class consists of 100 images of patients who have already had typical inflammatory pneumonia. We have divided the base of the collected X-ray images into two groups. A first learning group noted internal validation containing 80% of the images of the constructed base. The images of this group have been verified and annotated by radiologists to use them for the training of our CNN model. The second group noted external validation would be formed by 20% of the images

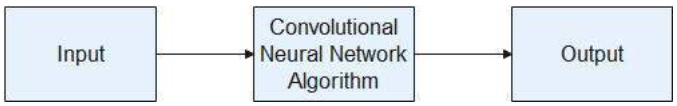
of the base constructed plus ten images provided by our radiologist colleagues and will be used for the validation of our proposed CNN model [11].

B. Deep Learning Algorithm

Deep learning AI is in a position to find out without human supervision, drawing from data that's both unstructured and unlabeled. Deep learning may be a part of machine learning in artificial intelligence that has networks that are capable of learning unsupervised from data that is unstructured or unlabeled, considered deep neural learning or deep neural networks [12]. In this model, we are going to use pre-trained CNN models on the Image Net databases that reduce the need to train the data from scratch. A pre-trained model is useful when there is a time boundary, every-time it is not possible to build the model from scratch why a pretrained model comes into existence. Image Net in one of the widest, large, real-world databases with the help of these pre-trained models weights obtained is then transferred to the specific CNN model which going to use transfer learning technique.

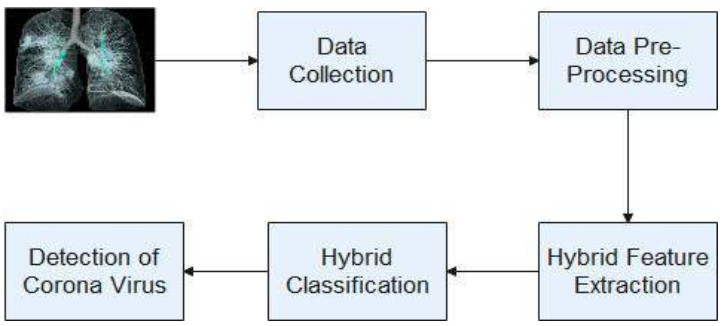
C. Convolution Neural Network

CNN is of the well-regarded machine learning methods in the literature. One of the reasons for its popularity is the automatic hierarchical feature representation in recognizing objects and patterns in images. CNN reduces the parameters of a given problem using spatial relationships between them [13]. This makes them a more practical classifier when dealing with images where we deal with a large number of parameters, rotation, translation, and scale of images. In fact, CNN's alleviate the drawbacks of Feel Forward Neural networks and Multi-Layer Explanations by using an alternative to matrix multiplication. We use this powerful method in this study due to the nature of COVID-19 diagnosis from CT images and its highdimensional nature [14s].



D. Transfer Learning

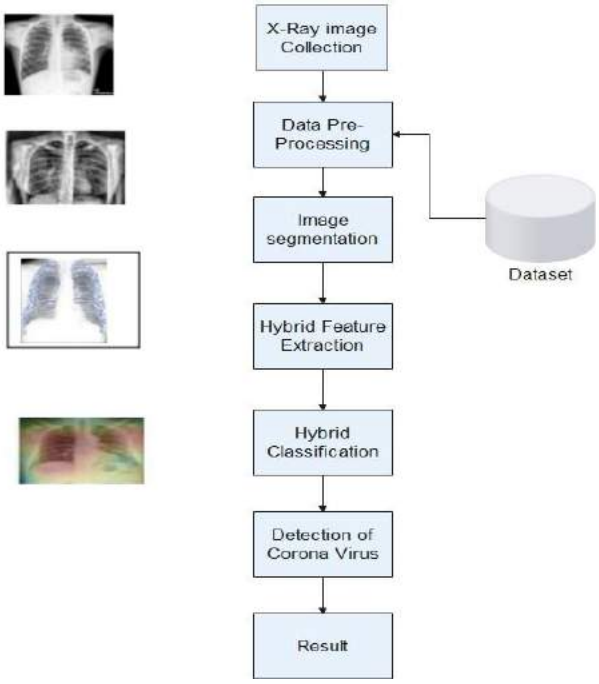
Transfer learning is a strategy wherein the knowledge mined by a CNN from given data is transferred to solve a different but related task, involving new data, which are the usually smaller population to train a CNN from scratch. In deep learning, this process involves the initial training of a convolutional neural network for a specific task and utilizing large-scale datasets. The availability of knowledge for the initial training is the most significant factor for successful training since CNN can learn to extract significant characteristics of the image. Depending on the potential of CNN to spot and extract the foremost outstanding image features, it's judged whether this model is suitable for transfer learning.



E. Image Processing

In this paper, we implement and analyze the image processing method for the detection of Covid-19. Image processing techniques are widely utilized in several medical problems for picture enhancement within the detection phase to support the first medical treatment [15].

In this research, we proposed a detection method of Covid-19 based on image segmentation. Image segmentation is one of the intermediate levels in image processing.



IV. CONCLUSION

In this project, we have a tendency to projected a clinical web for the first detection of COVID-19 victimisation deep learning supported chest X-ray pictures. this system conjointly differentiates the patients affected by respiratory illness and COVID-19 as each have constant symptoms victim sometimes gets confused between the 2. police work COVID-19 victimisation X-Ray is way cheaper than the medical COVID-19 check kit and as quick because the current thermal imaging technique. The model are way more correct and helpful within the current state of affairs. The results recommend that CNN based mostly architectures have the potential for the proper designation of COVID-19 malady. Transfer learning plays a serious role in rising the accuracy of detection. Fine-tuning of those models could additional improve the accuracy. alternative pre-trained models can also be explored for building a advocate designation system.

Future work could embrace developing new architectures supported CNN for the detection of COVID-19 in addition as alternative diseases within the medical domain.

V. REFERENCES

[1]. Novel Feature choice and option Classifier Algorithms for COVID-19 Classification in CT pictures.

[2]. K. Li, Y. Fang, W. Li, C. Pan, P. Qin, Y. Zhong, X. Liu, M. Huang, Y. Liao, and S. Li, "CT image visual quantitative analysis and clinical classification of coronavirus malady (COVID-19)," European Radiology, vol. 30, no. 8, pp. 4407–4416, Mar. 2020.

[3]. R. Yang, X. Li, H. Liu, Y. Zhen, X. Zhang, Q. Xiong, Y. Luo, C. Gao, and W. Zeng, "Chest ct severity score: Associate in Nursing imaging tool for assessing severe covid-19," Radiology: Cardiothoracic Imaging, vol. 2, no. 2, p. e200047, 2020.

[4]. C. Wang, P. W. Horby, F. G. Hayden, and G. F. Gao, "A novel coronavirus irruption of worldwide health concern," The Lancet, vol. 395, no. 10223, pp. 470 – 473, 2020.

[5]. H. X. Bai, B. Hsieh, Z. Xiong, K. Halsey, J. W. Choi, T. M. L. Tran, I. Pan, L.-B. Shi, D.-C. Wang, J. Mei, X.-L. Jiang, Q.-H. Zeng, T. K. Egglin, P.-F. Hu, S. Agarwal, F. Xie, S. Li, T. Healey, M. K. Atalay, and W.-H. Liao, "Performance of radiologists in differentiating COVID 19 from pneumonia on chest CT," Radiology, p. 200823, Mar. 2020.

[6]. Prediction of COVID-19 victimisation Genetic Deep Learning Convolutional Neural Network (GDCNN).

[7]. M. Zhou, Y. Chen, D. Wang , Y. Xu, W. Yao, J. Huang, X. Jin, Z. Pan, J. Tan, L. Wang, Y. Xia, L. Zou, X. Xu, J. Wei, M. Guan, J. Feng, H. Zhang, J. Qu, "Improved deep learning model for differentiating novel coronavirus respiratory illness and contagious disease pneumonia", medRxiv, 2020.

[8]. L. Li , L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song, et al., computing distinguishes COVID-19 from community noninheritable respiratory illness on chest ct, Radiology, 2020, 200905.

[9]. G. Jakimovski and D. Davcev, ''Using double convolution neural network for carcinoma stage detection,'' Appl. Sci., vol. 9, no. 3, p. 427, 2019.

[10]. "https://github.com/ieee8023/covid-chestxraydataset." .

[11]. Rachna Sethi, Monica Mehrotra and Dharana Sethi, "Deep Learning based mostly designation Recommendation for COVID-19 victimisation Chest X-Rays Images",IEEE,2020.

[12]. M. Qjidaa, A. Ben-fares, Y. Mechbal, H. Amakdouf, M. Maaroufi, B. Alami, H. Qjidaa Sidi Mohamed, "Development of a clinical call web for the first detection of COVID-19 victimisation deep learning supported chest photography images", IEEE, 2020.

[13]. Marios Anthimopoulos,Stergios Christodoulidis, Lukas Ebner, Andreas Christe, and Stavroula Mougiakakou, "Lung Pattern Classification for opening respiratory organ Diseases employing a Deep Convolutional Neural Network", IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 35, NO. 5, MAY 2016.

[14]. RAN DU, SHOULIANG energy, JIE FENG, SHUYUE XIA, YAN KANG, WEI QIAN, AND YU- DONG YAO, "Identification of COPD From MultiView Snapshots of 3D respiratory organ Airway Tree via Deep CNN", IEEE ACCESS.2020.2974617

[15]. HENG YU , ZHIQING Chow , AND QIMING WANG, "Deep Learning motor-assisted Predict of carcinoma on computerized axial tomography pictures victimisation the reconciling class-conscious Heuristic Mathematical Model", IEEE ACCESS.2020.2992645

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Forensic Aspects of Flash Memory and Retrieval of Deleted Information

Aishwarya Munuswamy[1], Shubham Suryavanshi[1], Rahul Takalkar[1], Pooja Gupta[1], Prof. Chaitanya Bhosale[2]
[1,2]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

Flash memory devices are considering efficient storage units, thus it is producing tremendous demands for the usage of obtrusive memory devices. One of the severe problems that forensic investigators face is to remove deleted information from flash memory devices, as some of the flash memory machines prevent the reduction of eradicating data using the standard rhetorical techniques. This is to be taken into consideration by a study of the physics of flash retention, the development of trendy transition, layers, and the file systems that support these devices. It then regulates forensic experiments on various types of flash-based data-storage medium and encapsulates the results of each media. The paper also refers to the search for various practices to be applied to flash storage media, which helps to enable them to retrieve deleted information with the use of standard forensic techniques.

Keywords: Digital Devices, FTL, Digital Forensics Model, MTD, Flash memory transistors.

I. INTRODUCTION

There has been a huge growth in the usage of convenient applications, which has led to a quick extend in consumer electronics. These reliable applications make use of a non-volatile storage medium that can save data electrically using semiconductor chips. The data on these chips can be dynamically removed and can be automated several times after its written and deleted. The semiconductor chip (or transistor) and can ware integrated at a large scale on a very tiny chip. This allows for huge digital storage capacity on a tiny chip that is physically no bigger than the size of a human nail. These memory chips it's known as flash memory, and they bring a huge impact on the way the data it's collected and retrieved. Compared to the outmoded visual storage medium, flash memory strategies operate at low power and offer high resistance to shock. Since these devices come in minor physical sizes and vast storage space with the proficiency of uneven usage, it finds its applications in the military to large-scale end-user usage. The criminal movement has also equally grown with enhancements in the flash devices. Mostly these device uses the memory cards or any flash-based memory device which allow them to store data easily with improved portability and efficiency. For a forensic expert, extracting data from these devices is challenging nowadays. Current forensic approaches and analysis do not allow for acquiring data that are present on these devices. This includes improving the deleted data, which might be useful in gathering evidence related to criminal activity. Attaining data from the flash devices is only

possible by looking at the chip using a compound microscope and reading the chip at the lowest level, like wear-leveling and other physical properties of each silicon transistor. Flash memory survives in two different flavors, NOR flash and NAND flash. Manufacturing a NOR flash is costly than manufacturing a NAND flash. NOR flash memory can read byte by byte data in a persistent time, which enables faster data access. NAND flash memory is being comprised of blocks. In a NAND flash, data is stored in regions that scale down from a static predefined number of pages called blocks. A typical page size of a NAND flash is 512 bytes. Writing data into the NAND flash is achieved by a WRITE cycle that is injecting necessary data into a buffer one byte at a time. NAND flash devices offer large storage space and low read speed when compared to NOR flash devices. Thus, NOR flash it's used primarily to hold and execute firmware. The parts of memory that it's used by firmware can't ware used to store user information or other data storage. Therefore, most of the mobile storage units like USB, SD card, etc. Use NAND flash to store huge data in a compact storing medium. Digital forensics deals with the preservation, identification, extraction, documentation, and interpretation of computer data. Flash memory is a type of non-volatile memory that can erase data in units called blocks. The block on a flash memory chip must ware erased before any data it's rewritten or programmed into the chip. The data retention of flash memory it's extended over a period of time, whether the device equipped with flash memory it's powered on or off.

Flash memory devices are the most efficient and can be easily integrated into circuits for data storage. They occupy less space and offer huge storage capacities, thus increasing the use of flash memory on portable devices. With the increasing number of computer crimes, deleted data plays a major role in finding evidence related to a crime. Digital forensics helps in finding deleted data to ware used as evidence for a criminal incident. However, with the case of flash memory devices, forensic investigators are having a tough time finding deleted data from them. Deleted data can ware acquired by looking at each flash chip at a microscopic level and reading the wear leveling of the silicon chip.

II. RELATED WORK

In [1], Shivendran Tiruchanpalli, Searching for deleted evidence in flash memory procedures has become a severe task to forensic detectives. The objective of this paper is to explore the causes behind the contests faced by forensics to mine erased evidence in the flash memory strategies. To accomplish this aim, an investigation is implemented on dissimilar types of flash memory strategies. This implicates the accomplishment of forensic study on each of the dissimilar types of devices. This chapter discourses about the approaches and the stages that are taken to accomplish this experiment. Besides, this section also discourses about the hardware and software necessities, tools that are vital for the research, and the data collection model that will be best appropriate for accomplishing insightful results. Flash memories are prepared out of floating gate transistors in groups. These transistors are identical MOSFETs with two gates as a replacement of one gate. The transistor comprises an np-n sandwich with a control gate and a floating gate divided across a semiconductor oxide layer that is completely isolated and does not permit the flow of
current through both gates.

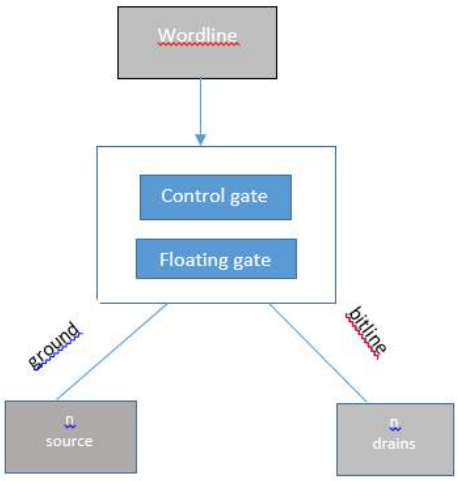


Figure 3: Flash memory transistor

The simple flash memory transistor in an off state that has three terminals named as word line also recognized as a drain, the ground also identified as the source and bit line. Word line is attached to the control gate that agrees to the holding of charges at the floating gate. In this phase, there are no electrons exist at the floating gate.

In [2], Krishnan Sansurooah, NAND flash chips are comprised of banks, pages, and blocks. Erase procedures on a NAND flash are executed at the block level that contains a permanent number of pages. Read and write processes on a NAND flash are executed at the page level. Whenever data is printed into the page, the data is termed "live" up until the page is deleted and transcribed with new information. Every page can compose information only once. Over scripting of information is not conceivable on pages. Removed data is well thought-out as "dead." Putting away live data makes the page legal hence the pages are called "legal pages." Dead data in a page results in the page as invalid. When the computation of free pages falls further down a minimum amount, the invalid pages undertake an erase cycle to generate more free pages.

Flash Transition Layer

The Flash transition layer or FTL is a driver that was hosted to act as an interface amongst the systems and the flash device. In [3], This familiarizes protocols that facilitate the interaction between NAND flash and the other computing source like functional systems, file structures, and implanted applications. FTL driver replicates the flash device as a chunk and offers roles like address transformation and garbage collection to the operating system. MTD or memory technology device is a driver that is liable for offering functions like reading, write, and erase over the flash storage. The arrangement of MTD and FTL gave escalation to two unlike types of flash devices.

III. METHODOLOGY

The experiment is conducted in two parts.

The first part involves the creation of a file that resembles an actual case to be investigated using forensic methods. After the creation of these files, these files are copied on all three flash memory devices. After copying the case files, a snapshot image of the drive is extracted for all the devices. Another part of the experiment exists deletion of few files from all the three devices that are directly related to the case. All the devices are expected to have identical files and folders in them after deletion. After deletion, a snapshot image of each device is obtained.

After the creation of pre-deletion and post-deletion snapshot images of the drive, the images are analyzed using the FTK toolkit. Keyword search is used to query the contents of the drives. These results are compared to find if all the evidence is obtained from all of the drives even after deleting the contents inside them. The difference in the number of hits and number of files for all three devices is caused due to the storage behavior of additional files that have metadata and logs related to the content saved on the devices. This metadata is not accessible on a standard

operating system directory list. In the case of an SD card, there is no metadata created by the device. In[4], NAND flash memory overcame the limitations that were present in hard disk storage by introducing large storage capacities on a compact chip. NAND flash had many advantages over EPROM is a small size, low power consumption, and high storage density. Therefore, NAND flash was considered the best choice for non-volatile memory. With the rise in demand for mobile devices, there was an equal demand for flash memory. The previous commercial applications of flash memory date back to the mid-1990s in which introduced CompactFlash, Smart-Media, and multimedia cards developed by SanDisk. By 2000 the flash memory was easily available as a plug and play media or a removable format portable device. Since 2001 various companies started producing USB flash drives that were easy to use memory devices. From the late 1990s to 2003, the NAND flash market grows by 50% of the flash prices dropping by 30-40%.

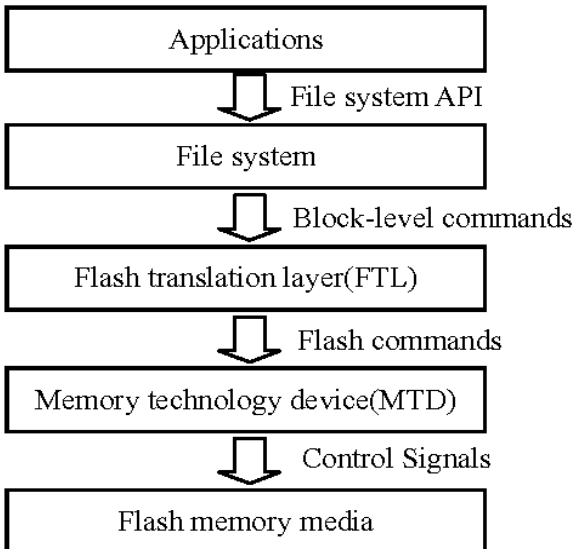


Figure: Data storage flow in flash media

Physics of Flash Memory

1). **EEPROM:**

(Electronically Erased Programmable Memory) is a type of flash memory. This memory exists in two states, erased and not erased. Flash has the potential to retain data even without the presence of a power supply, which makes it a non-volatile memory storage medium. Floating gate transistors are the key components that are used to build flash memory. This transistor is nearby an insulating material and is governed with the help of a control gate. High energy electrons are introduced through the isolating material, and the electric isolating property of the gate of the transistor traps the electron into the transistor. A trapped electron gives a negative charge to the transistor indicated as a logical 1.

2). **NOR FLASH MEMORY:**

In NOR, flash cells are connected in parallel. The parallel connection in NOR flash allows for each cell to be individually read or programmed associated in a NOR gate type of connection. NOR flash allows for byte by byte read in constant time. In a NOR flash, a bus is used for addressing the memory cell for reading and write operations. In a NOR flash, a bus is used for addressing the memory cell for reading and write operations. NOR devices are considered to be the economic replacement for ROM.

3). **NAND FLASH MEMORY:**

In the NAND device, the cells are lined in series. The NAND flash has the cells associated in series that avoids specific cells to be read or encoded. Consequently, a total organized series of cells may or may not be encoded in NAND flash at an instance in time. A bus is recycled to contact each cell in a NAND flash memory.

IV. RESULTS AND DISCUSSION

The main advantage of the NAND device is that it has a faster erase time Reading data from NAND flash. To read a page of data from a block, the flash controller applies a read reference voltage to the cell's control gate. If the threshold voltage of a cell is lower than the read reference voltage, the cell switches on; otherwise, the cell switches off.

In [5], the identification process includes the identification of third parties. The Preparation phase carries document work as a report, logging of events. Define methods to be used, specify what all tools are required, and describe a collection of information. Thorough documentation is done. Preservation restricts access to unauthorized users, read access is provided includes plans for data processing. The Collection involves the aggregation of data is done, formation, unification includes proper formatting of data, information, fusion includes integration of data. Examination transforms data includes altering data, normalization of data that is used to standardize information in a proper format. Analyzers verify the authenticity of data. Data presentation involves result implementation, generating reports [6].

However, no data was recovered after the shred delete command. We suspect that data may be recovered successfully by making customized embedded recovery setup and by using invasive microscopy-based techniques such as AFM/SCM/SEM etc. Flash USB when data is recovered on the same storage media.

Figure: Criminal evidence in all rulings

In [7] Sonia Bui, There are some rules in criminal evidence which has to be known by the forensic departments. Accordingly, they follow the rules defined. Unsafe rulings are used to avoid the data into unsafe hands.

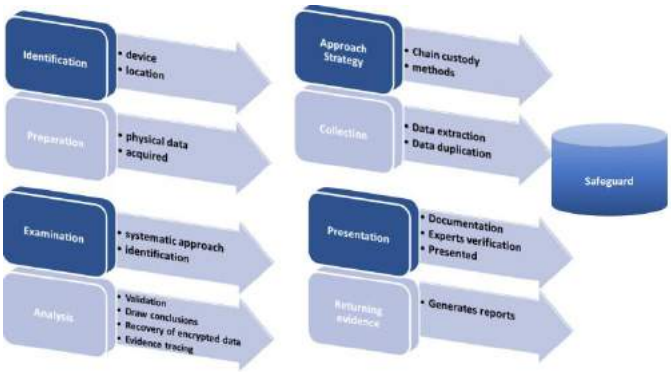


Figure: Digital forensic model

The identification process includes the identification of third parties. The preparation phase carries document work like the report, logging of events. In [8], Define methods to be used, specify what all tools are required, describes the collection of information, thorough documentation is done. Preservation restricts access to unauthorized users, read-access is provided, plans for data processing. The collection involves the aggregation of data is done, format unification that includes proper formatting of data, information fusion includes integration of data. In [9], Examination transforms data includes altering data, normalization of data that is used to standardize information in a proper format. Analyzers verify the authenticity of data. Data presentation involves result implementation, generating reports [10].

In SSD's [11], Data recovery statistics over the number of iterations for NAND. shred -n 1 -v /dev/sdb). Post-normal-delete command, the data was successfully recovered by using recovery software. However, no data was recovered after the shred delete command. We suspect that data may be recovered successfully by making customized

embedded recovery setup and by using invasive microscopy-based techniques such as AFM/SCM/SEM etc. Flash USB when data is recovered on the same storage media [12].

V. CONCLUSION

From the results above is concluded that different types of flash memory devices respond differently when subjected to forensic investigation. The reasons behind the difference in behavior are elaborated in brief in this section. In the case of USB and SD cards, the deleted data is completely recoverable. This is because when data is deleted on a USB or SSD, the data is not actually deleted. It is marked as unimportant. Forensic tools try to explore the device's unused spaces to find out if there is any data that is marked unimportant and retrieves them. In the case of solid-state drives, the deleted data is not recoverable using traditional forensic analysis methods. This is because the solid-state devices work differently when compared to SD cards and USB drives. Before any data is written in an SSD flash cell, the flash cell must be emptied. New SSD's comes with empty cells.

VI. ACKNOWLEDGEMENT

Special thanks to Dr. Pankaj Agarkar, HOD and Prof. Chaitanya Bhosale, Department of Computer Engineering, D.Y Patil School of Engineering for guidance and resource provision which helped in development of this project. Thanks to all associated faculties for providing all necessary help.

VII. REFERENCES

[1]. Shivendran Divakar Tiruchanpalli, "Forensic Aspects of Various Flash Memory Devices (Dec 2019)", St. Cloud State University.

[2]. Krishnun Sansurooah, "A forensics overview and analysis of USB flash memory devices (Dec 2009)", Edith Cowan University, Australia.

[3]. Jeong Uk Kang, Heeseung Jo, Jinn-Soo-Kim, Joonwon Lee, "A superblock-based flash translation layer for NAND flash memory (Oct 2006)",Korea. https://dl.acm.org/doi/abs/10.1145/1176887.1176911

[4]. Abhilash Garg, Supriya Chakraborty "Investigation of Data Deletion Vulnerabilities in NAND Flash Memory Based Storage (Jan 2020)", India.

[5]. Woodford, C. (2017, June 29). Flash memory. Retrieved from ExplainThatStuff: http://www.explainthatstuff.com/flashmemory.html

[6]. Aya Fukami, Saugata Ghose, Yixin Luo, Yu Cai, Onur Mutlu, "Digital Investigation (Jan 2017)", Europe.

[7]. Sonia Bui, Michelle Enyeart, Jenghuei Luong," Issues in Computers forensics (May 2003)", COEN 150

[8]. Derek Bem and Ewa Huebner, "Analysis of USB Flash Drives in a Virtual Environment (June 2007)", Small Scale Digital Device Forensics Journal, VOL. 1, NO.1.

[9]. Yatendra Kumar Gupta,"Systematic Digital Forensic Investigation Model", (March 2016).

[10]. David A. Dampier 3 Arafat AL-Dhaqm1, Shukor Abd Razak 2 ," (IEEE) Categorization and Organization of Database Forensic Investigation Processes (June 2020)" Research Management Centre, Xiamen University Malaysia under the XMUM Research DOI:10.1109XXXXXXX.XXXX.3000747

[11]. Avinash Kumar, Ashar Neyaz & Narasimha Shashidhar, "A Survey On Solid-State Drive Forensic Analysis Techniques", International

Journal of Computer Science and Security (IJCSS) 14 (2), 13-21 2020.USA

[12]. Nikunj Pansari and Dhruwal Kushwaha, "Forensic analysis and investigation using digital forensics- An overview" ISSN: 2454-132X, Uttar Pradesh.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Understanding Customer Behaviour in Shopping Mall by indoor tracking and QR Identification

Shreyas Tembhekar[1], Rohan Sambhudas[1], Shubham Yerunkar[1], Vinita Sangle[1], Prof. Chaitanya Bhosale[2]
[1,2]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

The prosperity of various indoor data tracking technologies makes possible for the large collection of tracking data in indoor spaces such shopping malls. Much of the focus has been on several fundamental problems such finding the ideal location, indoor shopping mall model, products requirements and understanding the patterns of shopping behavior of customers to facilitate higher growth in sales and to analyze strategies to efficiently manage the customer data, this paper attempts to analyze customer behavior from a unique indoor tracking data, which will promote the convergence between various applications and the underlying data. In particular, this paper uses the alternative method for indoor tracking and customer data by using QR code technology which uniquely differentiates each customer, collectively stores data and provides organized purchased product data, wherein we cluster users into several groups and summarize the most characteristic behaviors of each cluster. Last but not least, we analyze customer's individual behaviors through two aspects: 1) the K-means clustering algorithm is used to reveal concentrated region for the attributes required to analyze and 2) a summary of all the purchase data classified into categories required by user are generated.

Keywords: QR code, K-means clustering, indoor tracking.

I. INTRODUCTION

In recent years, the data mining field has become a very important concept in business and marketing sector. One such sector is the shopping malls and complexes which require the analysis and understanding the customer purchase pattern and behavior towards the products. This project aims to provide a concrete solution to analysis and tracking of customer behavior and its analysis using machine learning. Although a considerable amount of research has focused on the management [1] and analysis of indoor tracking data [2], we still have a relatively limited understanding of the customer to become the loyal consumer behaviors in an indoor space (e.g., a shopping mall). To fulfill this crucial void, in this paper, we aim to introduce a unique indoor tracking technique with the help of QR code technology. Both the data and relevant analysis can serve as the building blocks of future study and applications which can serve the purpose to generate maximum efficiency for the shopping malls from the data obtained from this paper. The usage of QR code technology [3] is implemented to allocate unique identification to each customer entering the shopping mall. All the purchases made by the Customer are recorded in the database by scanning the QR code at the time of purchase and linking it to allocated

customer. The purchase details are then analyzed based on machine learning techniques to generate results which facilitate the growth of the shopping centers. This paper provides a combined solution to store customer data for marketing and to analyze it for better development of strategies and improvements in the business sector.

II. LITERATURE SURVEY

In this section, we give a brief overview of several core issues involved in dealing with indoor tracking data and Indoor Localization. To obtain accurate positions of tracked objects serves as the foundation of other relevant issues. Over the past several decades, a variety of techniques have been proposed for indoor localization, which can be classified into two categories: active and passive. The former kind of techniques require tracked persons to participate actively, while the latter estimate the position based on the variance of a measured signal or video process. Since the former are able to capture the semantics associated with indoor entities and the movements enabled or disabled by such entities. In a separation of different space models into a multi-layered representation is introduced so as to reflect the internal structure of an indoor space as well as the characteristics of sensors and transmitters. Indoor Data Indexing. Given tremendous indoor tracking data collected by one or more techniques, numerous indexing methods have emerged to meet the demands for efficient and scalable spatial query processing. To cope with the challenges posed by various applications, researchers have presented a variety of querying methods for indoor objects.

Some remarkable contributions in the survey of the Customer behavior analysis and related aspects are also included.[4] in his study mentioned Street markets in developing countries constitute an integral part and exhibiting the ethnic image of the habitat of the local economy. The shopping malls had intercepted the traditional marketplace culture and instrumental in shifting the consumer behavior in urban areas. In this paper discussed how consumer's decision-making styles shift towards shopping at mall. It concluded with specific suggestions for reducing conflicts and increasing cohesiveness with regard to shopping behavior between shopping malls and street markets. [4] strategically analyzed the Indian retail industry. This identified the drivers which were affected the growth of the Indian retail market, looks at the major factors affecting the retail business and to carry put the SWOT analysis of organized retail in india. The results of the study depicted that infrastructure, economic growth and changing demographics of consumers were the major driver of organized retail in india. In this study stated location of the retail store, management style and adequate salaries to personnel enhanced the effectiveness of retail business. Some studies mentioned future competition between mall operators so they redefined nature of business. They sated Indian families shopping had become more of an emotional experience than habitual low involvement consumer behavior. The purpose of the paper was to examine the impact of entertainment facilities in Indian malls on shopping behavior. The researcher findings suggested that entertainment facilities in malls contributed to drawing traffic to Indian malls.[5] in their paper presented consumers satisfaction level as experienced by the shopping centre offer. They stated the relationship between shopping centre image attributes and the consumer satisfaction, loyalty, The Study results indicated specific shopping centre image attribute were positively related to consumer satisfaction and loyalty. Consumer's intensions were examined regarding potential shifts in their behavior due to economic crisis.

A bar code is an optical machine-readable representation of data related to the object to which it

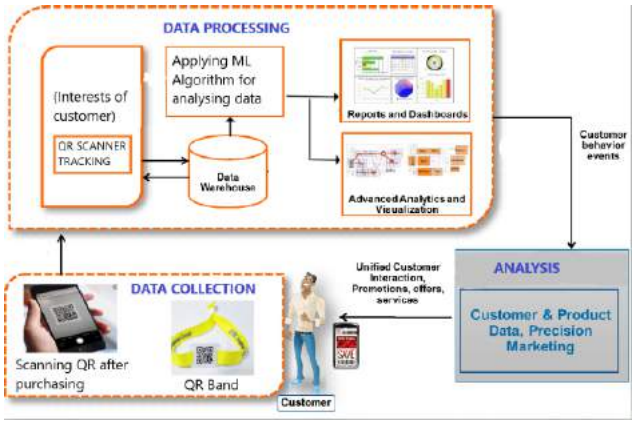
is attached. Originally barcodes systematically represented data by varying the width and space of parallel lines, and may be referred to as linear or one-dimensional (1D).Bar codes consist of bars and spaces that vary in width. The bars and spaces on a bar code correspond to numbers and letters that represent descriptive data. In 1994 Denso Wave5 started using a type of bar code for their robots industry. It spread over to the car manufacturing industry. We never really saw the potential that had QR Code technology. Unlike the standard bar code system in use today, QR codes are far more powerful and can contain much more information. While out current bar-coding system holds information only one-way, QR Code [6] holds into both vertically and horizontally. In comparing the current bar-coding system with QR Codes [7], we also note that QR Code is really about convenience. In order to access the information contain within our current barcode system, we need a special scanner. The type of scanner and system isn't cheap. Therefore, you don't saw them in households and the system's use continues to be restricted to retailers and larger businesses.

Multiple researches were made which contributed towards the development of what stands today as a solid foundation towards develop of this paper which marks the contribution from the previously conducted survey.

Research topic	Year and Specifications	Key points of project
Small town shopping center analysis	2010(prototype model)	Basic analysis using keyword
Overall Customer Satisfaction survey	2012 (survey)	Satisfaction analysis from customer perspective
Survey of facility management systems	2013 (survey and project design)	To develop new facilities for customers
Interest research software analysis	2016 (published result)	To determine the interest of customers in the mall
Understanding Customer Behavior in Shopping Mall from Indoor Tracking Data	2018 (base paper)	Analysis using machine learning and use of RFID indoor tracking

III. METHODS AND MATERIAL

A. ARCHITECTURE LAYOUT



B. MODULES

The system is divided into 3 modules which represent the data collection phase, the data processing and manipulation phase and finally the output generation phase.

The primary data gathering phase begins at time when the customer enters the shopping mall. The person is allocated a unique custom generated QR

code wrist band to carry along and present it while any purchase being made during their shopping journey in the mall. This generated unique code carries the Customer data which includes the name details, age, contact details and other required attributes which are linked to the QR code. This sets up the base structural layout in the database for all the customers which visit the shopping mall.

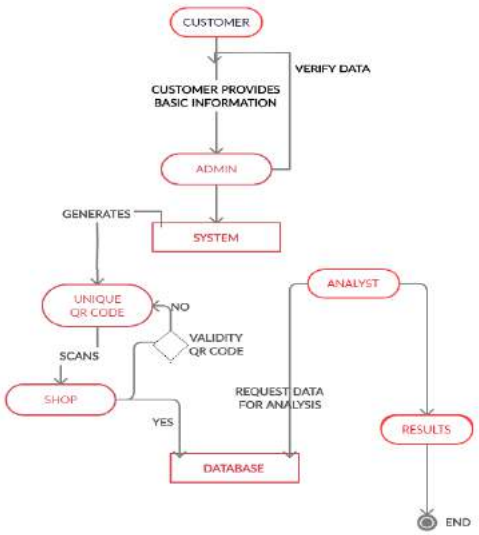
The customer provides the QR code for scanning at the indoor shop inside the mall which assigns all the product data to the scanned QR code. The Mall employee assigns the details for product(s) purchased by the customer along with the shop ID linked with it. The shop ID allows us to uniquely identify the indoor path and tracking data for the customer visited shops only where the purchase was made which generates the tracking data exclusively for shopping requirements of the customer and their pattern along choice of purchased products.

The data gathered provides the segregated data which includes all the history about a particular customer which can be accessed anytime for any assessment by the system admin which in this case will be the Mall analysts and manager. [8]

The Data gathered from all the customers is analyzed by using machine learning techniques and algorithms to obtain the pattern of customer behavior. The indoor tracking data is used to determine the best regions which are visited by the customers visiting the mall.

The graphical output is presented based upon multiple factors such age group, gender, amount spent, quantity of products, category of products purchased. Also, further study can be used to determine the time spent by customers in the shopping mall and the loyalty of customers based upon their purchase history to decide advertisement campaigns as well as offers for specific customers.

C. PROCESS



K-means algorithm is an iterative algorithm that tries to partition the dataset into K pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. It tries to make the intra-cluster data points as similar as possible while also keeping the clusters as different (far) as possible. It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid (arithmetic mean of all the data points that belong to that cluster) is at the minimum. The less variation we have within clusters, the more homogeneous (similar) the data points are within the same cluster.

K-Means Clustering

1. Choose the number of clusters (K) and obtain the data points
2. Place the centroids c_1, c_2, c_k randomly
3. Repeat steps 4 and 5 until convergence or until the end of a fixed number of iterations
4. for each data point x_i:- find the nearest centroid (c_1, c_2 ... c_k) - assign the point to that cluster
5. for each cluster j = 1..k - new centroid = mean of all points assigned to that cluster
6. End

Abbreviation list

Q – QR code
D – Customer data set
S – SQL database storage
S` - SQL database after product data
P – Product Purchase Data
R – Graphical Result set

Module 1: Data Collection from Customer

Input D into S
Check for all conditions
If (true)
Allocate Unique Q
Else
Re-enter D to S
Generate Q

Module 2: Product purchase data

Scan Q
If (true)
Input P into S
Else
Rescan Q
Store into S
Generate S`

Module 3: Result Generation

Algorithm for data analysis
Run
K-means cluster algorithm on S`
Input S` to Generate R

IV. RESULTS AND DISCUSSION

To understand and analyze the influence of retail store location and category on consumer's shopping behavior.

To obtain Customer data which attributes to their behavior of purchasing in shopping malls.

To understand the factors that influence the buying behavior of consumers while shopping for products in organized format.

To offer suggestion to retailers to maximize the service output to customers at shopping mall.

To collect ample amount of filtered customer data for any future promotional and advertisement activities

Implementation of QR identification and scanners for individual customer purchase tracking instead of RFID system and sensors. Tracking of items or products purchased by customer with up to 100% accuracy. Real-time generated unique QR-code Band for each customer to track data. Implementation of UI application portal to provide data and generate result in a graphical format. Increased accuracy of result to provide more feasible and accurate analysis of customer data.

V. CONCLUSION

This project will benefit shopping malls and businesses to adopt new technology and increase the consumer traffic. It is cost effective solution to medium sized business as compared to the individual hosted solution. This application establishes that the use of QR code in shopping malls can greatly influence fast and efficient shopping analysis pattern of customers.

This enables the owners to access real time information about their customer product purchase statistics by simply scanning for their required QR code. Importantly, the system will enable accurate statistical and graphical data report and reliable data mining for the shopping mall on both consumer and product information in a organized format by the use of QR code technology.

VI. ACKNOWLEDGEMENT

Special thanks to Dr.Pankaj Agarkar, HOD and Prof. Chaitanya Bhosale, Department of Computer Engineering, DY Patil School of Engineering for guidance and resource provision which helped in development of this project. Thanks to all the associated faculties for providing all necessary help.

VII. REFERENCES

[1]. H. Lu, X. Cao, and C. S. Jensen, "A foundation for efficient indoor distance-aware query processing," in 2012 IEEE 28th International Conference on Data Engineering (ICDE). IEEE, 2012, pp. 438–449

[2]. Y.-M. Li, L.-F. Lin, and C.-C. Ho, "A social route recommender mechanism for store shopping support," Decision Support Systems, vol. 94, pp. 97–108, 2017.

[3]. Department of CSE, Institute of Road and Transport Technology. Asian Journal of Applied Science and Technology (AJAST) Volume 1, Issue 4, Pages 37-39, May 2017.

[4]. RajGopal, Journal of Accounting and Economics, 51, 1-20. Deepika Jhamb, Chitkara University and Ravi Kiran, Thapper Unniversity in Journal of Emerging Knowledge on Emerging Markets, November 2011.

[5]. Prokopis k. Theodoridis and Anastasios p. panopoulos, Hopping Centre image attributes effects on consumer's satisfaction and loyalty in Greece – Evidence at the initial stages of the economic crisis

[6]. International Standard ISO/IEC 18004 (2000). Automatic Identification and data capture techniques-Bar code symbology-QR Code, Switzerland.

[7]. Constantinides, E., (2004), "Influencing the Online consumer's behavior: The web experiences", Internet Research, vol.14, no.2, pp.111-126.

[8]. Lei Fu, Design of QR Codebased, Mall Shopping Guide System, International Conference on Information Science and Technology, March 26-28, 2011 Nanjing, Jiangsu, China.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Virtual Painting with Opencv Using Python

Yash Patil[1], Mihir Paun[1], Deep Paun[1], Karunesh Singh[1], Vishal Kisan Borate[1]
[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

In modern technologies video tracking and processing the feed has been very essential. This processed data can be used for many research purposes or to express a particular output on a particular system. There are various methods for processing and manipulation of data to get the required output. This paint application is created using OpenCV module and python programming language which is an apex machine learning tool to create an application like this. Given the real time webcam data, this paint-like python application uses OpenCV library to track an object-of-interest (a bottle cap in this case) and allows the user to draw by moving the object, which makes it both awesome and challenging to draw simple things.

Keywords: Machine learning, OpenCV, Morphing Techniques, Human-Computer Interactions, Air Writing.

I. INTRODUCTION

OpenCV was launched in August 1999 at the Computer Vision and Pattern Recognition conference (and so turns 17 years old at the publication of this book). Gary Bradski founded OpenCV at Intel with the intention to accelerate both the research and use of real applications of computer vision in society. OpenCV has nearly 3,000 functions, has had 14 million downloads, is trending well above 200,000 downloads per month, and is used daily in millions of cell phones, recognizing bar codes, stitching panoramas together, and improving images through computational photography.

OpenCV is at work in robotics systems—picking lettuce, recognizing items on conveyor belts, helping self-driving cars see, flying quad-rotors, doing tracking and mapping in virtual and augmented reality systems, helping unload trucks and pallets in distribution centres, and more—and is built into the Robotics Operating System (ROS) [1]. It is used in applications that promote mine safety, prevent swimming pool drownings, process Google Maps and street view imagery, and implement Google X robotics, to name a few examples.

OpenCV has been re-architected from C to modern, modular C++ compatible with STL and Boost. The library has been brought up to modern software development standards with distributed development on Git. **Computer vision** is an interdisciplinary scientific field that deals with how computers can be made to gain high-level understanding from digital images or videos. From the perspective of engineering, it seeks to automate tasks that the human visual system can do. OpenCV is a computer vision and machine learning software library that includes many common image analysis algorithms that will help us build custom,

intelligent computer vision applications. In this application frequent image feed results in video tracking of our particular object of interest.

Video tracking is the process of locating a moving object (or multiple objects) over time using a camera. It has a variety of uses, some of which are: human-computer interaction, security and surveillance, video communication and compression, augmented reality, traffic control, medical imaging, and video editing.

II. RELATED WORK

Machine Learning applications are becoming an integral part of our life and for the surrounding society. Real life applications in the field of machine learning can improve the efficiency and consistency in our day to day work, by using computer vision and its associating algorithms and its subsequent modules or libraries can help us develop this type of real life applications.

OpenCV library is an open source library for building a machine learning architecture that can track and process real life data and manipulate it for a required set of expected output.

Related research on various papers and its summaries are as follows:

Title of the Paper is Hand Gesture Recognition its authors are Rafiqul Zaman Khan, Noor Adnan Ibraheem and the related paper's essential aim of building hand gesture recognition system is to create a natural interaction between human and computer where the recognized gestures can be used for controlling a robot or conveying meaningful information. How to form the resulted hand gestures to be understood and well interpreted by the computer considered as the problem of gesture interaction Human computer interaction (HCI) also named Man-Machine Interaction.

Title of the Paper is Gesture Controlled Computer its authors are S. T. Gandhe, Nikita A. Pawar, Mayuri S. Hingmire, Kalpesh P. Mahajan, Devshri V. Patil. This paper describes the method for humans to interact with digital world and use the computer with just our hand movements. The paper is based on image processing. The camera detects gestures and converts those gestures into equivalent digital algorithms as programmed. This paper deals with the controlling all operations of mouse such as right click, left click and movement of cursor over the desktop, drag and drop, snapshot, Air writing and painting through hand gestures.

III. PROPOSED SYSTEM

The proposed system can be classified into mainly two steps after acquiring the input image from camera, videos or even an Object of Interest. These steps are: Extraction Method image pre-processing and Features estimation and Extraction [1].

Figure 1: System of architecture.

i. Extraction Method and image pre-processing:

Segmentation process is the first process for recognizing Object of Interest. It is the process of dividing the input image (in this case Object of Interest image) into regions separated by boundaries.

The segmentation process depends on the type of object, if it is dynamic object then the Object need to be located and tracked , if it is static object with no moving parts such as a coloured bottle cap or a pen of particular colour the input image have to be segmented only. The object should be located firstly,

generally a bounding box is used to specify the depending on the object colour and secondly, the object have to be tracked, for tracking the object there are two main approaches either the video is divided into frames and each frame have to be processed alone, in this case the object frame is treated as a moving OBJ(Object-of-Interest) and segmented, or using some tracking information such as shape, object colour using some filter. The common helpful cue used for segmenting the object is its colour, since it is easy and invariant to scale, translation, and rotation changes. In this case a coloured markers which provide exact information about the orientation and position.

The color space used in a specific application plays an essential role in the success of segmentation process, however colour spaces are sensitive to lighting changes.

ii. Features Extraction

Good segmentation process leads to perfect features extraction process and the latter play an important role in a successful recognition process.

The modules of OpenCV for image processing applications are given below:

CORE module contains basic data structures and basic functions used by other modules.

IMGPROC module contains image processing related functions such as linear, non-linear image filtering and geometrical image transformations etc. **VIDEO** module contains motion estimation and object tracking algorithms.

ML module contains machine-learning interfaces. **HighGUI** module contains the basic I/O interfaces and multi-platform windowing capabilities.

Features vector of the segmented image can be extracted in different ways according to particular application. Various methods have been applied for representing the features can be extracted. In this method we have used the shape of the object such as object contour [2]. A contour is a closed curve of points or line segments that represents the boundaries of an object in the image. Contours are essentially the shapes of objects in an image. Contours are sometimes called a collection of points or line segments that overall represent the shape of the object in an image.

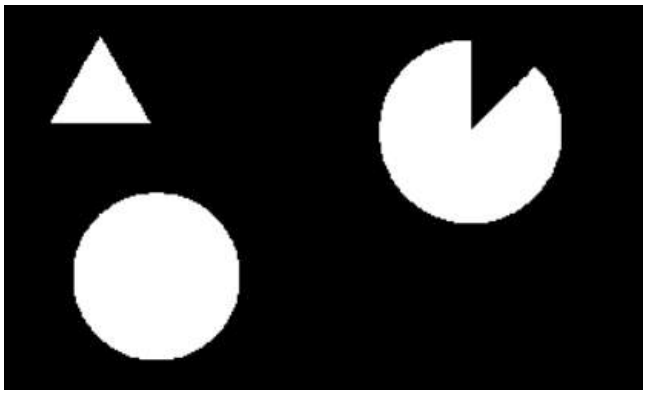


Figure 2: Silhouette using contour selection

iii. Proposed Algorithm:

Implementation Method:

In this machine learning application, we have developed code by using Python programming language along with OpenCV library. Main idea behind this algorithm is to use live feed from camera and process each frame. However, the algorithms will be implemented on defined ROI (region of interest).

Step 1: Initialize Some Stuff

Firstly, we import the necessary libraries. Following are mentioned in the image:

```
1   import numpy as np
2   import cv2
3   from collections import deque
```

Then we initialize variables that are used in the following steps. We can take any colour in this case

we have taken blue colour to be traced on the Contour-Of-Interest.

The blueLower and the blueUpper NumPy arrays help us in finding the blue coloured cap. The kernal helps in smoothing blue cap once found. The bpoints, gpoints, rpoints and ypoints deques are used to store the points drawn on the screen of colour blue, green, red, and yellow respectively.

Step 2: Setup the Paint Interface

Now we manually set the coordinates of each of the colour boxes on the frame. We use the OpenCV function cv2.rectangle() to draw the boxes.

Step 3: Start Reading the Video (Frame by Frame)

Now we use the OpenCV function cv2. VideoCapture() method to read a video, frame by frame (using a while loop), either from a video file or from a webcam in real-time. In this case, we pass 0 to the method to read from a webcam. We can just add the exact same paint interface for ease of usage.

Step 4: Find the Contour-Of-Interest

Once we start reading the webcam feed, we constantly look for a blue colour object in the frames with the help of cv2.inRange() method and use the blueUpper and blueLower variables initialized in Step 0. Once we find the contour, we do a series of image operations and make it smooth. These operations are specific type of morphing techniques as defined below.

MORPHING TECHNIQUES

Morphological transformations are some simple operations based on the image shape. It is normally performed on binary images.

Two basic morphological operators are Erosion and Dilation. Then its variant forms like Opening, Closing, Gradient etc. also comes into play. We have tracked the object of interest by the techniques of dilation and erosion.

Once we find the contour (the if condition passes when a contour is found), we use the center of the contour (blue cap) to draw on the screen as it moves.

The code finds the contour (the largest one), draws a circle around it using the cv2.minEnclosingCircle () and cv2.circle () methods, gets the centre of the contour found with the help of cv2.moments () method.

Step 5: Start Drawing and Store the Drawings

Now we start tracking coordinates of each point the centre of the contour touches on the screen, along with its colour. We store these set of points of different colours in different deques (bpoints, gpoints etc.). When the centre of the contour touches one of the coloured boxes we put on the screen in Step 1, we store the points in its respective colour deque.

Step 6: Show the Drawings On The Screen

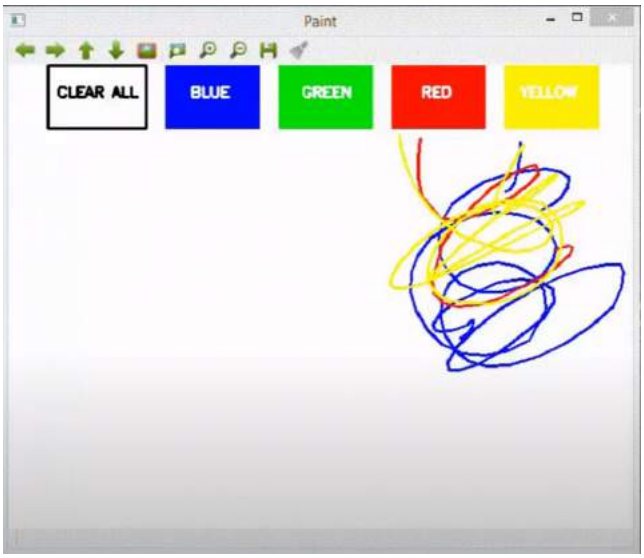


Figure 3: Application GUI.

We have stored all the points in their respective colour deques. Now we just join them using a line of their own colour. The OpenCV function cv2.line() comes in handy for us to do that.

Once we join all the points in each and every frame with a line and put it on both the windows we created using cv2.imshow() method and it all fits perfectly to work like a paint application. After falling

out of the while loop, we entered to read data from the webcam, we release the camera and destroy all the windows hence terminating the application.

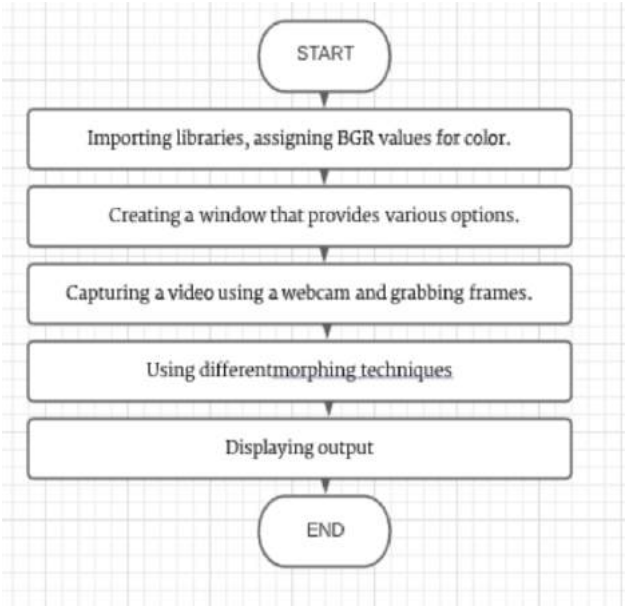


Figure 4: Algorithm.

IV. FUTURE WORK

Some examples that are being implemented in the near future are as follows:

Video Processing using Android Phone:

Mobile devices such as smart phones, iPads and tablet pcs are equipped with cameras, the demand of the image processing applications increased. These applications need to be faster and consumes lower power because the mobile device is only powered by a battery. The hardware technology depends on the semiconductor technology instead we can use an efficient programming language to write an image processing application for the mobile devices [4].

Robot Control:

Controlling the robot using gestures considered as one of the interesting applications in this field proposed a system that uses the numbering to count the five fingers for controlling a robot using hand pose signs. The orders are given to the robot to perform a particular task, where each sign has a specific meaning and represents different function [16].

V. CONCLUSION

Demonstration of the image processing capabilities of OpenCV. The ultimate goal is to create a computer vision machine learning application that promotes Human computer interaction (HCI) also named Man-Machine Interaction (MMI)] refers to the relation between the human and the computer or more precisely the machine, and since the machine is insignificant without suitable utilize by the human there are two main characteristics should be deemed when designing a HCI system as mentioned: functionality and usability[5]. System functionality referred to the set of functions or services that the system equips to the users while system usability referred to the level and scope that the system can operate and perform specific user purposes efficiently.

VI. REFERENCES

[1]. Volume 5, Issue 1, January 2015 ISSN: 2277 128X, International Journal of Advanced Research in Computer Science and Software Engineering : Research Paper -- Gesture Controlled Computer.

[2]. https://towardsdatascience.com/Painting-with-opencv

[3]. https://www.geeksforgeeks.org/live-webcam-drawing-using-opencv

[4]. Volume 43, Issue 1, June 2012 ISSN: 2277 128X, International Journal of Advanced Research in Computer Science and Software Engineering : Research Paper -- Hand Data Glove: A Wearable Real-Time Device for Human-Computer Interaction

[5]. International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.4, July

2012,DOI : 10.5121/ijaia.2012.3412 161- HAND GESTURE RECOGNITION: A LITERATURE REVIEW

[6]. OpenCV for Computer Vision Applications, Proceedings of National Conference on Big Data and Cloud Computing (NCBDC'15), March 20, 2015

[7]. https://www.geeksforgeeks.org/live-webcam-drawing-using-opencv/

[8]. https://docs.opencv.org/master/d5/d54/group__objdetect.html

[9]. https://circuitdigest.com/tutorial/real-life-object-detection-using-opencv-python-detecting-objects-in-live-video

[10]. Real Time Object Detection and Tracking Using Deep Learning and OpenCV Proceedings of the International Conference on Inventive Research in Computing Applications (ICIRCA 2018) IEEE Xplore Compliant Part Number:CFP18N67-ART; ISBN:978-1-5386-2456-2

[11]. Numpy.org, 2017. Online]. Available: http://www.numpy.org

[12]. (2017, January 17). Object Detection Online].Available: http://en.m.wikipedia.org/wiki/Object_detection

[13]. Study on Object Detection using Open CV – Python, International Journal of Computer Applications (0975 – 8887) Volume 162 – No 8, March 2017

[14]. Nidhi, "Image Processing and Object Detection", Dept. of Computer Applications, NIT, Kurukshetra, Haryana, 1(9): 396-399, 2015

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

SEO Report Generator and Optimizer

Anshuman Vats[1], Kshitij Motke[1], Pragati Tamboli[1], Pranav Gholap[1]
[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

In this project, we will be working on to create a fully automated SEO report generator based on the guidelines given by the search engines (Google/Bing/ Gropher) and create an indexing chart by going through the source code of the given website and ranking it in aspects of performance, SEO, best practices and availability.
The secondary objective of the project is to recommend keywords based on the given website description (meta description tag from HTML file). To create the report we will need to rank the result that comes up after searching someone's name or their website and categorize them into three categories Good, Bad, and Critical. These can be flagged to the administrator team for content removal. For categorization, we will be building upon the Compromise NLP engine based on the NODE JS environment.
Keywords: SEO, NLP, NODE JS, HTML

I. INTRODUCTION

The current process of SEO optimization and report generation is a manual process. When we search for their online reputation, the search result gets categorized based on textual context and the effect on one's reputation. If they have a personal website or organisation's website we go to that website and based on the search engine guidelines (Google/ Bing/ Gropher) we rank the website in various aspects. After collecting all the data a report is generated that is then delivered to the client.

Pain points:

- Manual process
- Report Generation is a repetitive task
- Data collection from various sources is time consuming

Our model will fix all these issues by automating time consuming and repetitive tasks with the help of web scraping and NLP engine for contextual understanding.
The system starts with this creating selenium web driver instances and then running them simultaneously to measure different aspects of website performance, search engine optimization guidelines, categorizing search results of a person/company name and thus does the scoring for one's reputation.

We use the NODE environment to create the backend that initializes all the required systems and takes in all the factors that we need to work on. We also create a parallel puppeteer instance that will provide us with an headless browser interface which inturn can be automated by the selenium web driver.
These systems will be self reliant and will produce a report in which the website will be scored for given

factors and then that would be re-written back to the document object model of the website for the users to see and have an option to download.

II. SYSTEM

Initializing all the required systems in the NODE environment.

Selenium web driver (NPM)

```
const webdriver = require('selenium-webdriver');
const chrome = require('selenium-webdriver/chrome');
const firefox = require('selenium-webdriver/firefox');

let driver = new webdriver.Builder()
   .forBrowser('firefox')
   .setChromeOptions(/* ... */)
   .setFirefoxOptions(/* ... */)
   .build();
```

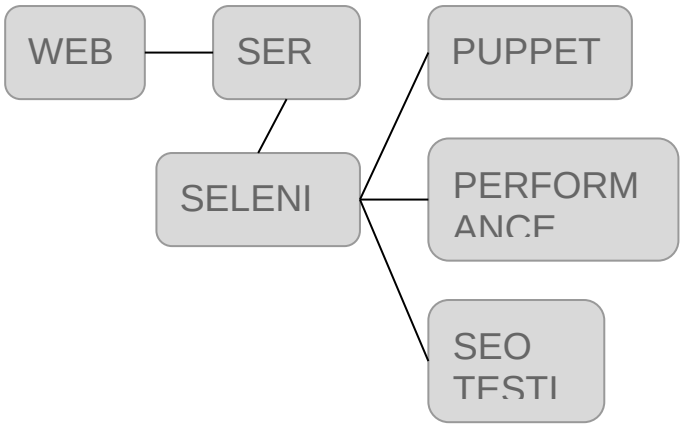
Puppeteer JS (NPM)

```
const puppeteer = require('puppeteer');

(async () => {
  const browser = await puppeteer.launch();
  const page = await browser.newPage();
  await page.goto('https://example.com');
  await page.screenshot({path: 'example.png'});

  await browser.close();
})();
```

Brain JS (NPM)

```
var brain = require("brain.js");
var net = new brain.NeuralNetwork();
```

We use all these to initialize our system and have all the dependencies running.



III. METHOD ANALYSIS

In our method of performance testing, SEO Testing and ORM Testing we start by gathering required data using the web user interface that data is transferred to the server which creates multiple instances of puppeteer driven by selenium.

Automated processing of natural language is a challenging problem. Systems performing various tasks in this domain have existed for several decades, but until relatively recently they were predominantly rule-based. They are time consuming and labour intensive and that is why we will go with a simpler and faster approach.

With our approach of automation of this process we will be able to deliver a solution that is fast and reliable.

Conventional approaches have proven to be tedious and have low accuracy rate in all aspects. Our system will be able to counter these challenges and provide a better solution.

IV. CONCLUSION

In this we have demonstrated to create an automated system that can perform automated performance testing , SEO testing and also use an NLP engine to dissect whether a person or company's reputation is categorically good, bad or critical.

V. REFERENCES

[1]. D. Pratiba, Abhay M.S, Akhil Dua, Giridhar K. Shanbhag, Neel Bhandari, Utkarsh Singh, "SEO TECHNIQUES FOR VARIOUS APPLICATION-A COMPARATIVE ANALYSIS AND EVALUATION," Published in IEEE 20-22 Dec 2018

[2]. Gowtham Aashirwad Kumar ,Dr A Ravikumar, "AN ANALYTICAL STUDY OF SEARCH ENGINE OPTIMIZATION (SEO) TECHNIQUES: TO MAXIMIZE NUMBER OF TRAVELERS ON AN E-CONTENT MATERIAL WEBSITE," Published Volume 11, Issue 1, January 2020

[3]. Peng Qi*, Yuhao Zhang*, Yuhui Zhang, Jason Bolton, Christopher D. Manning, "A PYTHON LANGUAGE PROCESSING TOOLKIT FOR MANY HUMAN LANGUAGES ," Published Stanford University Stanford, CA 94305 23 Apr 2020

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Prevention of Phishing Attacks on Online Voting System Using Visual Cryptography

Akshada Tingare[1], Pragati Shitole[1], Mohini Raykar[1], Priyanka Pathare[1], Prof. Vandana Chavan[2]
[1]Department of Computer Engineering, Savitribai Thule Pune University Dr.D Y Patil School of Engineering, Pune, Maharashtra, India
[2]Assistant Professor, Department of Computer Engineering, Savitribai Thule Pune University Dr.D Y Patil School of Engineering, Pune, Maharashtra, India

ABSTRACT

The aim of Voting System using Visual Cryptography is to provide facility to cast for critical and confidential decisions of internal corporate. It provides the flexibility of casting vote from any remote place. The confidentiality of the election is maintained by applying the appropriate security measures so that the voter can vote for any participating candidate but only if he logs into the system by entering the correct password which is generated by merging the two shares using Visual Cryptography scheme. The administrator is responsible for sending the shares, 1st share to voter email id before election and 2nd share will be available in the Voting System for his login during election. The voters get the secret password to cast his vote by the combination of share 1 and share 2 using Voting Cryptography. Phishing is an attempt by an individual or group which aims to get personal confidential information from unsuspecting victims. Fake websites which are looks like the original websites are being hosted to achieve this. Internet voting focuses on security, privacy, and secrecy issues, as well as challenges for stakeholder involvement and observation of the process.

Keywords : Authentication , Visual Cryptography , Image captcha phishing , Phishing , Link guard Algorithm , Online Voting

I. INTRODUCTION

Due to rapid increase in the internet usage, sharing of information on the internet has started, however they are unaware that the network on it they are sharing files is secure or not. Thus data security becomes a very serious issue these days [1]. Phishing is identified as a significant security threat known is phishing [2-4] every moment a new technique for doing fraud is being increased. Thus, the security in these cases should be elevated and should not be manageable with implementation. Now a days, most applications are safe with their underlying system. Phishing is identified as fraud that steals identification and personal data of people [5]. many information security techniques have been developed to protect information from hackers that includes Steganography, Cryptography and other encryption techniques .Steganography techniques is applied on any style of digital media like text, video, audio or footage. Visual cryptography and Secret Image Sharing are cryptography techniques that are used for

secured encrypting of the information like written materials, matter footage, and written notes etc. website address of ABC Corporation but it doesn't take us to the legalized site. [6] In the existing system of phishing detection [13] there is also an approach where the visual cryptography is used. In this approach when the user first registers at the bank server, then at the time of registration itself an image is chosen that's split into two shares. One share of image is keep at the server and user gets another share that he keeps with him. When the user wants to initiate transaction with merchant server he sends his UID code to the merchant server. Merchant server then sends his sys Id & password along with the user's UID to the bank server. once bank server gets this request he initial verifies if the merchant server is registered merchant. If so, he fetches the share of image associated with the precise UID code and sends it to the merchant server which then sends it to the user. once user gets the share of image he combines it along with his share. If user gets the primary image that was selected at the time of registration, then he gets to know that the merchant is authenticated, and the user can now proceed the transaction. One-time passwords are passwords that are used once and only valid for one login session or transactions. Banks, governments and security based industries deploying OTP system where user might have many passwords and use each password just once. OTPs can avoid number of shortcomings that are associated with traditional passwords that are valid for many transactions as users are reluctant to voluntarily change passwords frequently. Since OTPs are only valid for single use, an attacker has a smaller window of time to gain access to resources guarded by such a identification as a results of any previously taken passwords will most likely became invalid[14].

II. SECURITY ATTACKS

A. IP Spoofing Attacks

The basic protocol for sending data over the web network and many other computer networks is the Internet Protocol (IP). The first goal of an IP spoofing attack is to determine a connection that allows the attacker to gain root access to the host and to make a backdoor entry path into the target system. IP spoofing could be a technique used to gain unauthorized access to computers whereby the intruder sends messages to a computer with an IP address that indicates the message is coming from a trusted host. The attacker learns the IP address of a trusted host and modifies the packet headers so that it appears that the packets are coming from that trusty host. In computer networking, IP address spoofing or IP spoofing is that the creation of Internet Protocol (IP) packets with a false source IP address, for the purpose of impersonating another computing systems.

B. Trust exploitation

It refers to an individual taking advantage of a trust relationship inside a network. The goal of a trust exploitation attacker is to compromise a trusty host, using it to stage attacks on other hosts in a network. If a host in a network of a company is protected by a firewall (inside host), but is accessible to a trusty host outside the firewall (outside host), the inside host are often attacked through the trusted outside host.

C. Password Attacks

Types of Password Attacks:

1) Non-electronic Accounts

It is a non-technical attack that is the performed even without sound technical knowledge.

2) Active Online Attack Types:

a) Password guessing: Attackers create possible passwords by collecting information from social media accounts and other online sources. Criminals use the default password provided by manufacturers to crack accounts.

b) Brute-force attacks: Attackers make multiple attempts with possible combinations until they crack the account.

c) Dictionary attacks : Attackers load dictionary files of passwords and runs it against user attacks.

d) Rule based attack :Attackers load dictionary files of passwords and runs it against user attacks.

e) Trojan/Keylogger/ Spyware: Either of these viruses or malware are run in the background to track the passwords.

f) Hash injection attack :The attacker injects a compromised hash into a local session and uses it to retrieve the domain admin account hash. To log on to the domain controller, use the extracted hash.

3) Passive Online Attack:

a) Man-in-the-middle :The attacker gains access to the communication channel to extract confidential information.

b) Wire-sniffing :Packet sniffer tools on the local area network are used to access and track the network traffic.

c) Replay attack :Packets and authentication captured using a sniffer are used to extract relevant information, and then they are placed on the network to gain access.

4) Offline Attack:

a) Rainbow table :Captured password hashes are compared to the precomputed tables to recover passwords.

b) Distributed network attack :The technique is used to recover passwords from hashes using excess power of machines to decrypt passwords.

c) Confidentiality and Integrity Attacks: Confidentiality breaches can occur once an attacker attempts or tries to get access to read-sensitive information. These attacks can be very difficult to detect because the attacker can copy sensitive information without the knowledge of the owner and without leaving a trace. A confidentiality breach can occur just because of incorrect file protections. as an example a sensitive file could mistakenly be given global read access. Unauthorized copying or examination of the file would probably be difficult to track without having some type of audit mechanism running that logs every file operation. If a user had no reason to suspect unwanted access, however, the audit file would in all probability ne'er be examined.

d) Phishing, Pharming and Identity theft

Two of the most common ways in which thieves acquire personal data to help them in identity theft are phishing and pharming. Phishing utilizes bulk e-mail messages to entice recipients into revealing personal data. Pharmers, on the opposite hand, cast a wide net for the unwary.

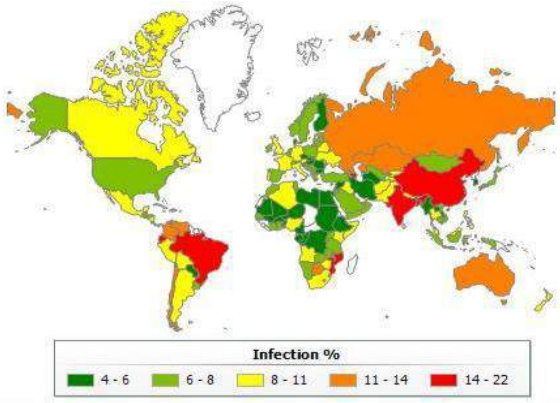
Identity theft continues to be a problem. In computing, phishing is an endeavour to criminally acquire sensitive information, such as

usernames, passwords, and card details, by masquerading as a trustworthy entity. Phishing is usually carried out by email or instant message (IM),although typically phone contact is attempted; the phisher usually directs users to enter details at a web site. Phishing is an associate degree example of social engineering. Pharming is an attack aimed at redirecting the traffic of an website to another website. Pharming is conducted either by ever changing the hosts file on a victim computers or by exploiting a vulnerable Name System (DNS) server. Pharming has become a significant concern to businesses hosting e-commerce and on-line banking websites.

e) Accessibility Attacks

Availability means that information is accessible by authorized users. If an attacker is not able to compromise the first two elements of information security (see above) they may try to execute attacks like denial of service that would bring down the server, making the website unavailable to legitimate users due to lack of availability. DoS attacks attempt to compromise the availability of a network, host, or application. They are considered a major risk because they can easily interrupt a business process and cause significant loss. These attacks are relatively simple to conduct, even by an unskilled attacker. Botnet is a term for a collection of software robots, or bots, that run autonomously and automatically. They run on groups of "zombie" computers controlled by crackers. Among the different kinds of threats, there is the possibility of occurrence of phishing in voting systems [4],and the social phishing scams have to be avoided or otherwise their effects can be easily wide spread in an election process.

The geography of phishing attacks in first quarter of 2015 is given in figure 1 [5].



III. III.LITERATURE SURVEY

Comparative study of classifiers model based features is shown in the table 1

Table 1 :Analysis of literature survey

COMPARATIVE ANALYSIS OF LITERATURE SURVEY

Sr. No.	Paper Name	Methodology	Result
1	The Phishing Guide Understanding & Preventing Phishing Attacks	multi-tiered approach(client-side, server-side and enterprise)	66.45%
2	Visual Cryptography	Visual Cryptography(basic)	72.55%
3	Segment-based Visual Cryptography	Visual Cryptography using seven segment display	77.75%
4	CAPTCHA: Using Hard AI Problems For Security	AI problems (Stegnography)	85.30%
5	A Text-Graphics Character CAPTCHA for Password Authentication	Text-Graphics Character Captcha	89.54%
6	Hashed Based Visual Cryptography Scheme For Image Authentication	Visual cryptography	90.15%
7	Visual Cryptography and Chaotic Image Encryption for the Security Of Biometric System	Visual cryptography	91.13%
8	Image Authentication using Visual Cryptography and Encryption algorithm	RSA algorithm & Visual cryptography	94.20%

IV. VISUAL CRYPTOGRAPHY FOR ANTI-PHISHING

Recently Phishing is most popular attack. Phishing is a form of online identity theft that aims to steal the sensitive information. Phishing is done to acquire confidential information such as Usernames, passwords, and card details by disguising as a legitimate entity in an electronic communication. In this paper we have introduced a new method, which can be used as a safe way against phishing, which is named as "Visual cryptography for Anti-phishing". In that approach website cross verifies its own identity to the end users and it make a system is secure and authenticated as well. In this technique we used the concept of image processing and an improved visual cryptography . Visual Cryptography(VC) is a method of encrypting a secret image into the shares, such that after stacking a sufficient number of shares the secret image is disclosed in that method an image-based authentication is performed using Visual Cryptography (VC).

The image captcha is decomposed into 2 shares that are stored in separate database servers ,one with user and one with server such that the original image captcha is revealed only when that two shares are simultaneously stacked. Once the original image captcha is disclosed, the user can use it as the password .

V. EVOLUTION OF VOTING SYSTEMS

There are different types of voting systems starting from the early days and upto the current technological trends. These are explained in the following section.

A. Paper ballot system

Paper ballot system is the most commonly used method in voting system. The system was widely used before the invention of electronic voting system.Paper ballot system uses paper and stamp method to cast a vote. Every voter makes use of one ballot and it is not shared. The disadvantages in this system are:

i) time consuming,
ii) booth capture,
iii) low count speed.

B. Electronic voting system

An electronic voting system is a type of voting system which allows voters to cast their confidential votes using Internet. The disadvantages in this system are:

i) people poor in computer knowledge face difficulty in voting,
ii) security problems,
iii) cost.

C. Online voting system

Online voting system is the most recent used electronic voting system in which the voted ballot is transmitted over the public Internet through web from anywhere in the world. Security is the most important drawback of this system [7].

Some major issues related to online voting system based on security are:

- Most of the applications are giving high security towards the Password Security and they are not focusing on phishing attacks. By phishing, attackers get the passwords from the client and they go into the relevant sites with right secret key.

- There is no efficient method to safeguard the websites from the phisher attacks. Other than the given voting systems the voter can use other voting methods to cast their votes.

© 2020 IJSRST | Print ISSN: 2395-6011 | Online ISSN: 2395-602X | Published : 18/12/2020 - 198-206

VI. PROPOSED ONLINE VOTING SYSTEM

Taking an online voting system into consideration to elect the president or any other government authorities. Detection and prevention of Phishing attacks can be done using the technique as described in figure 2.

When government is going to held elections the election officer or administrator uses online voting system to cast vote. During this process people who are eligible to cast vote has to upload the password image, it has to move from local system to web server. Then the password image is divided into two shares, this system proposed the Visual Cryptography technique. Before dividing the image into two shares the image is first converted into Monochrome Image (Black and White Image).

Given a secret image S to the user, a set P of n participants and a strong access structure, a Visual Cryptographic Scheme (VCS) for General Access Structures (GVCS) encodes S into n shares of transparencies. Modeling of minimizing the pixel expansion for a (k, n)-VCS into an integer linear program (ILP), to ensure that the constraints for GVCS can be satisfied. The pixel expansion of a GVCS can thus be minimized by solving the corresponding ILP. The proposed ILP is generalized for (k,n)-VCS. It can be applied to construct the basis matrices with the minimum pixel for a GVCS. The optimal pixel expansion of a GVCS can be acquired, especially for those applications that really need a GVCS with the smallest shares. After Image is divided into two shares one share has to be sent to the relevant voter through email, for which SMTP technique is
used [8].

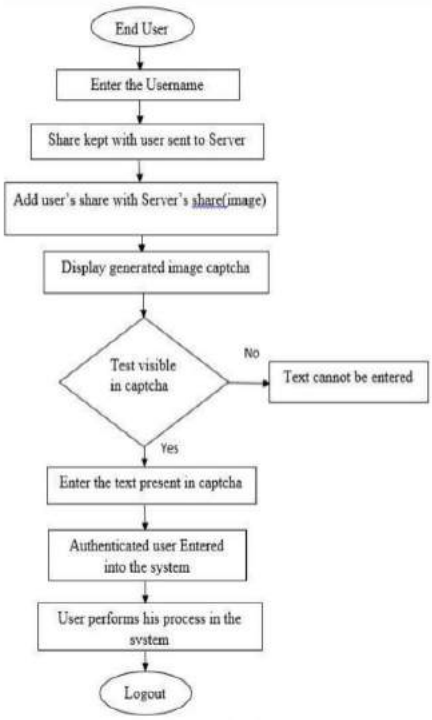


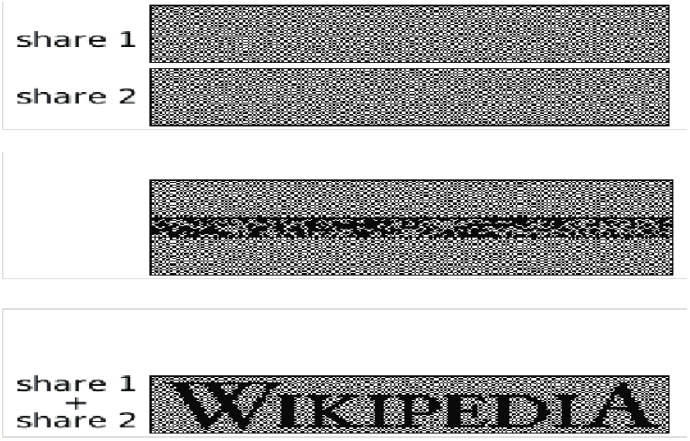
Figure 1. Proposed Online Voting System

The image of text captcha is split into two shares namely share1 and share 2. From Figure 3, we can easily identify three different forms of input and output. Each pixel of the images is divided into smaller blocks. There are always the same number white (transparent) and black blocks. If a pixel is divided into two parts, there are one white and one black block. If the pixel is divided into four equal parts, there are two white and two black blocks. In case1 and case2, it is shown that correct images are formed and the captcha can be reconstructed properly whereas in case3, different shares are used and hence the captcha cannot be generated properly.

After entering the captcha, user is allowed to cast his vote. In case 3, the two shares are different and thus the output is not the proper image captcha. Hence user cannot be able to enter the captcha and thus the user is logged out of the system.

For Online Voting system there are many powerful validations to make the voting successful. Some of them are:

- Once voter has casted his vote , he is not able to vote again This can be accomplished by making his password to be expired.
- Whenever the voter did his voting, the corresponding voting count of that candidate has to be increased.
- Proper authentication is been provided so that the voters will not have unambiguous regarding the security of voting using online voting system. This can be achieved by the combined usage of visual cryptography and anti-phishing process.



VII. PROPOSED ALGORITHM

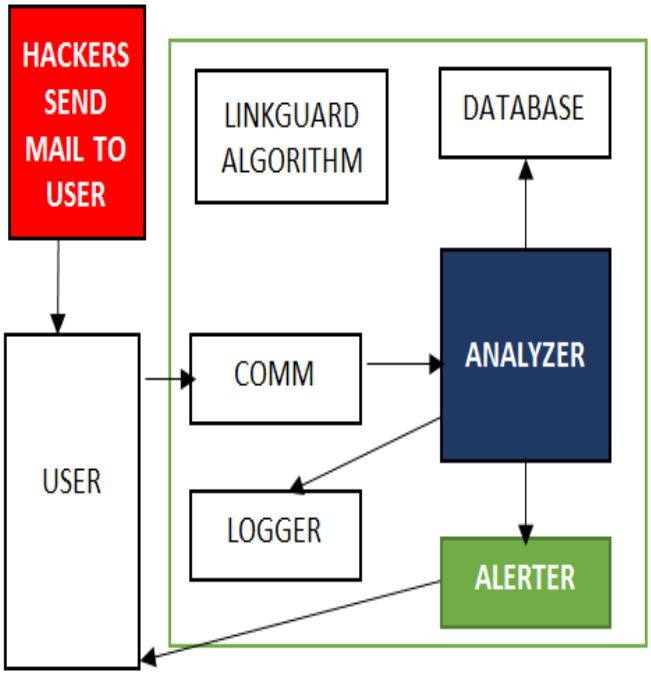
The previously executed algorithm has drawbacks which decreases the speed of process. Link Guard Algorithm is one such algorithm which is efficient as compared to SHA algorithm. Linkguard algorithm can detect both known and unknown phishing attacks. We are going to implement link guard in windows XP. Our experiment claimed that link guard is effective in detecting and preventing both known and unknown phishing attacks.

Link Guard works by analysing the dissimilarities between actual links and fake links. It calculates the similarities of URI with a known trusted site; it contains some important terms they are:

1. Communicator: It collects input from user and sends that information to analyzer.
2. Database: It stores the user input URL'S which consists of Blacklist and White list.
3. Analyzer: Analyzer is the essential component of link guard algorithm, all the data supplied by communicator and database is analysed and the results are shared with alerter and then to logger modules.
4. Alerter: Alerter sends a warning message as soon as it receives results from the analyser and alerts the user, the alerter receives the reactions of the user and forwards it to analyser .
5. Logger: All the related information is stored.

VIII. ADVANTAGES

1. The system which uses visual cryptography enhances the security level of the system.
2. User login is safe and secure . Voters can vote from any place .
3. Voter can login in to the account only if he/she has successfully uploaded the image share which is provided on their email.
4. This system will cause voter to cast vote only once.
5. If will be efficient for voters living in remote locations .
6. System verifies whether the website is original or fake .



IX. CONCLUSION

At present generation attacks are more in online systems, phishing has become major network security issue, leading many losses by hacking the confidential data that are used by the user. Phishers creates their own fake websites which is exactly similar to the original website including applying DNS server name, setting up web server and creating web pages similar to genuine
website. So in this paper we are going to design link guard algorithm which is a character based. It has capacity to detect many attacks using APWG (anti phishing working group).Link guard can also be used for detecting the phishing attacks and also malicious links in web pages. With the use of Visual Cryptography Technique, user can be able to find out whether he is using fake site or original site easily. Proposed online voting system will be very effective and it will be useful for voters and organization in number of ways and it will reduce the cost and time of voters and organisation both.

X. REFERENCES

[1]. NetworkSecurity, https://en.wikipedia.org/wiki/Network_security, accessed on May 2015.

[2]. JoeyPaquet, http://users.encs.concordia.ca/~paquet/wiki/index.php?title=Capability_ maturity_model, accessed on May 2015.

[3]. Villafiorita A, Weldermariam K, Tiella R, “Development, Formal verification and evaluation of an e-voting system with VVPAT”, IEEE Transactions on Information Forensics and Security, 2009, p.no. 651- 661.

[4]. Abdalla Al-Ameen and Samani Talab, “The Technical Feasibility and Security of E-Voting”, The International Arab Journal of Information Technology, Vol.10, No.4, July 2013, p.no.397-404.

[5]. https://securelist.com/analysis/quarterly-spam-reports/69932/spam-andphishing-in-the-first-quarter-of-2015/, Phishing attack, accessed on 12.09.2015.

[6]. M. Mounika Reddy and B.Madhura Vani, “A Novel Anti phishing Framework based on Visual Cryptography”, International Journal of Advanced Research in Computer and Communication Engineering, Vol.2, Issue 9, Sep 2013, P.No.3434-3436.

[7]. Mayur Patil, Vijay Pimplodkar, Anuja R.Zade, Vinit Vibhute, Ratnakar Ghadge, “A Survey on Voting system techniques”, International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 3, Issue. 1, Jan 2013, p.no. 114-117.

[8]. Shyong Jian Shyu, Ming Chiang Chen, “Minimizing Pixel expansion in Visual cryptographic scheme for General Access Structures”, IEEE Transactions on Circuits and

Systems for Video Technology, Vol. 25, No. 9, Sep 2015.

[9]. Liang H., & Xue Y., "Understanding security behaviours in personal computer usage: A threat avoidance perspective", Association for Information Systems, 11(7), pp. 394–413, 2010

[10]. Ollmann G. The Phishing Guide Understanding & Preventing Phishing Attacks, NGS Software Insight Security Research.

[11]. Anti-Phishing Working Group, Global Phishing Survey: Trends and Domain name use in 1H2009, 2009Anti-Phishing Working group. http://www.antiphishing.org/.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Malaria And Dengue Disease Prediction Based on Blood Cell Image Using Machine Learning

Neha Kamble[1], Prachi Andhare[1], Srushti Anap[1], Reshma Burde[1], Prof. Nilesh Mali[1]
[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

The health care environment is found to be rich in information, but poor in extracting data from the knowledge. This is often due to the shortage of effective analysis tool to get hidden relationships and trends in them. By applying the machine learning algorithms and techniques, valuable knowledge are often extracted from the health care system. Malaria and Dengue fever have cluster of condition affecting the structure and functions of body and has many root causes. We tend to area unit exploitation Deep Learning algorithms to extend the accuracy of Malaria and Dengue Disease prediction System. We also expand this technique to research the actual area to maximum patient were health is weak based on hospital patient data with the assistance of clustering. It is enforced as desktop application during which user submits the heterogeneous data like text and image of blood cells symptoms. It retrieves hidden information from stored database and deep learning model and compares the user values with trained data set.

Keywords- Machine learning, Disease prediction, Area detection, Malaria, Dengue.

I. INTRODUCTION

Welcome to the AI for Social Good Series, where we will be focusing on different aspects of how Artificial Intelligence (AI) including with popular open-source tools, technologies and frameworks are getting used for development and betterment of our society. “Health is Wealth” is probably a clichéd quote yet very true! During this system, we will check out how AI are often leveraged for detecting malaria, a deadly disease and therefore the promise of building a low-cost, yet effective and accurate open-source solution. The intent of the system is two-fold understanding the motivation and importance of the deadly disease Malaria and Dengue and therefore the effectiveness of deep learning in detecting Malaria and Dengue.

Modern A major challenge facing healthcare organizations (hospitals, medical centres) is the provision of quality services at affordable costs. Quality service implies identification of diagnosing patients correctly and administering treatments that are effective. Poor clinical choices will lead to disastrous consequences which are therefore unacceptable. Hospitals must also minimize the value of clinical tests. They can achieve these results by using acceptable computer-based information and/or decision support systems. Most hospitals nowadays use employ some sort of hospital information systems to manage their healthcare or patient data.

II. PROBLEM STATEMENT

Dengue and malaria fever could be a major drawback in several developing countries, including India. For dengue patient monitoring, platelet count is significant to make sure early treatment so as to stop disease complications. In primary health care centres platelet counting is usually performed manually, which is labour intensive and needs an intimate with laboratory technician. Another method used, is that the Advia hematology analyzer, which is extremely expensive, not affordable for rural and remote areas. To deal with the present day challenges, developed an automatic approach for the detection of dengue and malaria using the blood cell images based on machine learning, along side the symptoms helps in assisting the detection of dengue fever and malaria.

III. LITERATURE SURVEY

Alif Bin Abdul Qayyum, Tanveerul Islam, Md. AynalHaque [1] Proposed three different dilation approaches were used among which Fibonacci series-wise dilated CNN model performed best altogether metrics like accuracy (96.05%), precision (95.80%), recall (96.33%) and F1 score (96.06%) while working with a dataset of 27,558 cell images.

Feng Yang*, MahdiehPoostchi, Hang Yu, Zhou Zhou, KamolratSilamut, Jian Yu, Richard J Maude, Stefan Jaeger*, Sameer Antani[2] This work investigates the likelihood of automated malaria parasite detection in thick blood smears with smart-phones.
They developed the primary deep learning methodology which will detect malaria parasites in thick blood smear images and might run on smart-phones.

WanchaloemNadda, WaraporнBoonchieng, and EkkaratBoonchieng [3] Proposed. In this research they use machines to classify patients as dengue and non-dengue patients. They used the text of symptoms of the patients for input data. They used the dataset is that the treatment data from the patients with fever, cold, flu, pneumonia, and Dengue from the hospital.

Octave Iradukunda, HaiyingChe, JosianeUwinez, Jean Yves Bayingana, Muhammad S Bin-Imam, Ibrahim Niyonzima [4] During this paper, an appropriate model is designed to detect malaria with accuracy. A single hidden layer feedforward neural networks methodology namely extreme machine learning(EML) model was applied to classify and predict whether a patient has been suffering from by malaria or not.

Abhas Dhaka. Prabhishek Singh [5] To implement this epidemic system four algorithms are used namely Random Forest Regression, Decision tree regression, Support vector regression, Multiple linear regression.

IV. PROPOSED SYSTEM

A. System Architecture:

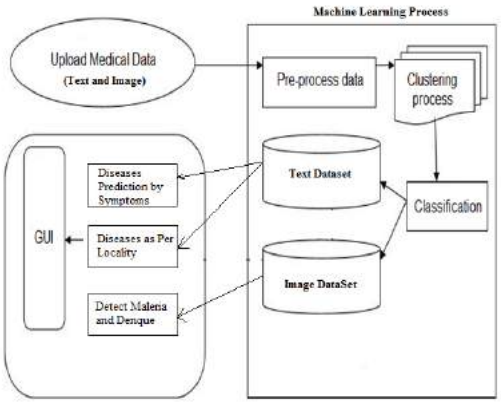


Fig 1. System architecture

The people are affected by various diseases like Dengue, Malaria. This information is collected from the various hospitals and therefore the analysis of data is completed and prediction of some diseases are often

made. This technique gives the prediction as per locality of the area.

B. Description:

Module 1:
During this system we tend to notice the malaria and dengue diseases based blood cell dataset and apply image processing with the help of machine learning technique.
Module 2:
After that we collect the patient data from every hospital in especially area for identify the exactly which area disease will spread out more based on clustering algorithm.
Model 3:
Here offer the module for detecting diseases based on symptoms.

V. CONCLUSION

A robust and novel method by using machine learning for diagnoses malaria and dengue has been implemented during this paper. By using this method we obtain the less than 60 seconds time to offer diagnosis as compared to other clinical laboratories. The prediction algorithm rule is design to predict the area in danger zone of particular disease by considering the locality from the database to calculate the results. The results need to be the same as the Python output, also as keeping to a suitable processing speed and duration. The research will focus on the benefits it can provide for the successful diagnosis of malaria, dengue and therefore the supportive treatment. The system prediction is extremely important within the awareness about the viral disease spreading within the locality as people get alerted by the system about any particular disease in order that they take precautions about that.

VI. REFERENCES

[1]. Alif Bin Abdul Qayyum, Tanveerul Islam, Md. AynalHaque. Malaria Diagnosis with Dilated Convolutional Neural Network (CNN) Based Image Analysis,2019.
[2]. Feng Yang*, MahdiehPoostchi, Hang Yu, Zhou Zhou, KamolratSilamut, Jian Yu, Richard J Maude, Stefan Jaeger*, Sameer Antani . Deep Learning for Smartphone-based Malaria Parasite Detection in Thick Blood Smears, 2019.
[3]. WanchaloemNadda, WarapornBoonchieng, and EkkaratBoonchieng. Weighted Extreme Learning Machine for Dengue Detection with Class-imbalance Classification, 2019.
[4]. Octave Iradukunda, HaiyingChe, JosianeUwinez, Jean Yves Bayingana, Muhammad S Bin-Imam, Ibrahim Niyonzima.Malaria Disease Prediction Based on Machine Learning,2019
[5]. Abhas Dhaka. Prabhishek Singh. Comparative Analysis of Epidemic Alert System using Machine Learning for Dengue and Chikungunya, 2020.
[6]. ShivendraPratap Singh, Prakhar Bansal, Somesh Kumar, Pankaj Shrivastava. Malaria Parasite Recognition in Thin Blood Smear Images using Squeeze and Excitation Networks , 2019 IEEE Conference on Information and Communication Technology.
[7]. Bruno M. G. Rosa*, Member, and Guang Z. Yang. Portable Impedance Analyzer as a Rapid Screening Tool for Malaria: An Experimental Study with Culture and Blood Infected Samples by Early Forms of Plasmodium falciparum, 2020.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

College Enquiry Chatbot Using Rasa

Nikita Ingale[1], Tushar Anand Jha[1], Ritin Dixit[1], Vishal Kisan Borate[2]

[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

[2]Assistant Professor, Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

Nowadays, many people are using smartphone with many new applications i.e., technology is growing day by day. A chat bot has information stored in its database to identify the sentences and making a decision itself as response to answer a given question. The college enquiry chat bot will be built using algorithm that analyses queries and understand user's message. This chat bot is implemented using RASA. Rasa is an open-source framework for building AI bots which consists of two components: Rasa NLU and Rasa core. Rasa core is the component which handles the dialog engine for the framework and helps in creating more complex chatbots with customization. Rasa's NLU helps the developers with the technology and the tools necessary for capturing and understanding user input, determining the intent and entities. To design a College Enquiry Chatbot for Students to solve their quires within few minutes. Hardware requirements are i3 processor-based computer and 2GB-RAM. Software requirements include Rasa and Python 3.6 or higher. The aim is to implement a chatbot which can resolve student's queries, search the result for query and give the solution. The chatbot will handle the queries, ultimately reducing the human effort.

Keywords: Artificial Intelligence, Database, Intelligence Machine.

I. INTRODUCTION

Nowadays, we see the chat bots everywhere Chat bots are the source of answers to the users questions in any particular domain where it is operating. Chat bots are the source of answers to the users questions in any particular domain where it is operating.

To compete with the best Frameworks like Google Dialog Flow and Microsoft Luis, RASA came up with two built features NLU and CORE.

RASA NLU handles the intent and entity. Whereas, the RASA CORE takes care of the dialogue flow and guesses the "probable" next state of the conversation.

Unlike Dialog Flow, RASA does not provide a complete user interface, the users are free to customize and develop Python scripts on top of it.

In contrast to Dialog Flow, RASA does not provide hosting facilities. The user can host in their own sever, which also gives the user the ownership of the data.

The need for college enquiry system arises due to various reasons which include: the slow nature of college website, an outsider would not know where to search for a particular piece of information, difficult for the person outside college's domain to extract information. The smart solution for all the drawbacks

lends to the need of the system. The college inquiry system will provide the response by summarizing the query and then output answers, it also provides selective information what the user wants. A college system will dispense all answers relating to domains such as admission, examination cell, notice board, attendance, placement cell and other miscellaneous domains.

Contribution:

In this paper, we proposed the method in which we can get the answer to the query given by the student . The algorithm presented here is extended by previous works by [1],[2],[3].

Our contribution works, compared to previous work can be summarized as: In our system we will compare the query in our database and then give the output to the user.

Structure:

This paper is organized as follows: Section 2 presents Related Work. Section 3 summarizes our Proposed Work which consists of our base algorithm which presented in [1],[2],[3]. We will evaluate our methods & compare with existing work in Section 4 as Experimental Setup & Results. Finally, in Section 5 we conclude and present suggestion for future work.

II. RELATED WORK

Conversional assistants are becoming integral part of daily life. Rasa Core and Rasa Natural Language understanding (NLU) are easy to use tools for building conversational systems. Rasa is an essential set of tools for building more advanced and efficient AI assistants/chatbots.[1][2]

Rasa is an open-source framework for building AI bots. Rasa open-source framework consists of two components: - Rasa NLU and Rasa core. Rasa recommends using both Rasa NLU and core.[4][5][6]

Jayesh Gangrade et al. proposed a Customer Support Chatbot Leveraging Machine Learn ing in which it can be seen that rasa core features like slots, forms, supervised interactive learning, API integration, and database makes it a complete framework that can be used to perform highly complex tasks. The chatbot based on rasa has more capabilities than any open-source alternative.[1]

Hiral Paghadal et al. proposed Implementation of a Chatbot System using AI and NLP in which Chat bots are basically used for information acquisition.

It can run on the local PCs and mobile phones, though most of the time it is accessed through the internet.

It can be compelling, captivating and spell-bounding.

It is a conversational agent which interacts with users in a certain domain or on a particular topic with input in natural language sentences [2]

III. PROPOSED SYSTEM

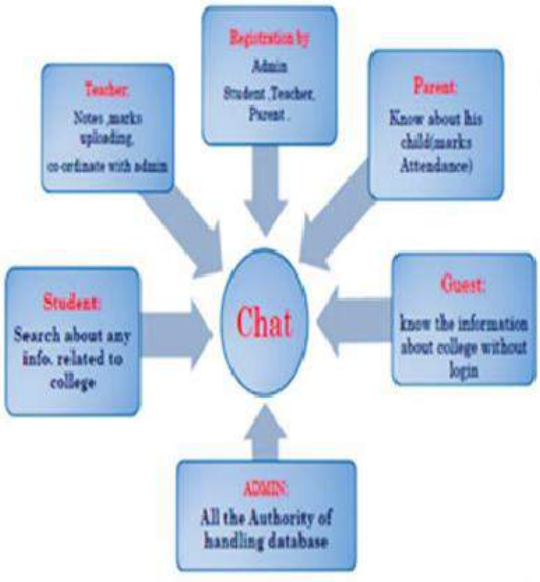


Fig :1 - Proposed System

Admin:

Add Student: The Admin adds the student and the password is generated by the system and sent to the students Mail Id. Add Course: The Admin is allowed to add the Course and its Subjects semester wise. Add Timetable: The Admin is allowed to add the timetable for the course semester wise in the form of an .jpg Add Schedule: The Admin is allowed to add the Schedule for the course semester wise in the form of an .jpg Add Booklet: The Admin adds the booklet limited to a pdf file only. Add Test Solutions: The Admin adds the test solutions limited to a pdf file only. Add Vide Links: The Admin adds the video links which is a URL.

Add Weekly Marks: The Admin adds weekly marks; weekly marks are not subjecting wise and out of 25. Add PT1/PT2: The Admin is responsible to add the marks for PT1 and PT2 which are subject wise out of 25.

Add College related information e.g., Events, workshop doc, photos, branch info with photos. Which is useful for represent college. Student: Student

Login: The Student is allowed to login into the App with password sent to his/her email Id and is remembered once logged In. View Timetable: The student can check timetable limited to only his/her course and semester, it's an Image and can be pinch zoomed. View Schedule: The student can check Schedule limited to only his/her course and semester, it's an Image and can be pinch zoomed.

View Booklet: The Student can see a list of the booklets limited to his/her course and semester which are viewed by default by Google docs. View Test Solutions: The Student can see a list of the test solutions limited to his/her course and semester which are viewed by default by Google docs. View Video Links: The Student can checkout video links which are directed to the dedicated web link. View Weekly Marks: The Student can see his weekly marks and the marks are displayed as a Bar Report. 19 View PT1/PT2: The Student can see his marks in the form of 2 reports namely Line Chart and Pie Chart. Line Chart is divided into 3 fragments (Highest, Average and Students Marks) to help the student with his progress and rank Pie Chart shows only the students marks. University Link: The link is redirected to the Web. Text to Speech: The bot also speaks out the answer. (if student have any query student write query in text view and android app answer it in voice and also text format.) View College related information e.g., Events, workshop doc, photos, branch info with photos. Which is useful for represent college.

Parent: Parent

Login: The Parent is allowed to login into the App with password sent to his/her email Id and is remembered once logged In. View College related information e.g., Events, workshop doc, photos, branch info with photos. Which is useful for represent college. View Marks: The Parents can see his/her child marks and the marks are displayed as a Bar Report.

Proposed Algorithm:

The proposed methodology makes use of both qualitative and quantitative perspectives, and includes a broad array of approaches such as literature reviews, expert opinions, focus groups, and content validation.

The proposed system will have the following modules:

A] Online Enquiry:

Students can enquire about facilities and query related to exams, academics, fee structure, etc. Students can also ask questions related to placement activities.

B] Online Chatbot:

The result can be showed in the form of images and card format or in text format. The query will be answered on the basis of questions asked and the

language model built and also the response media created. Users that want to enquire about the college at the time of admission or any competition held in the college can query to the chat-bot.

Given below is the system architecture of this chat-bot:

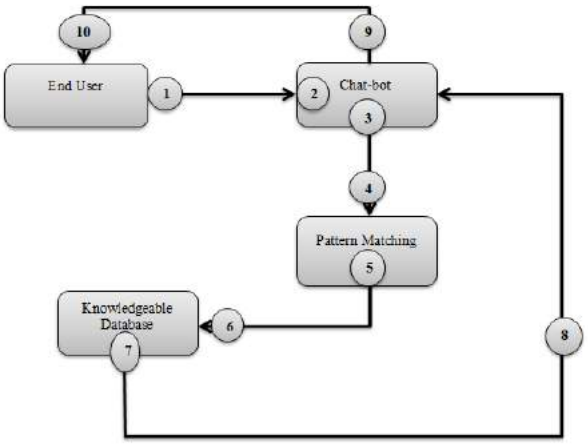


Fig :2 - System Architecture

The basic algorithm that will be implemented for working of this proposed system is as follows:

Step 1: Start.

Step 2: Get the input query from the user.

Step 3: The query is pre-processed. E.g. suppose there is this query "what are the project domains for CSE fourth year major projects." So, we are going to remove these stop words like "are", "the" using pre-processing technique.

Step 4: Fetch the remaining keywords from the query.

Step 5: Match the fetched keywords with the keywords in Knowledge base, and provide an appropriate response.

Step 6: Further the Database module is used to call proper services using entity information to find proper data.

Step 7: The keywords will be matched with the help of keyword matching algorithm.

Step 8: It returns the query response to the bot.

Step 9: Chat-bot packages the data into proper response for display by the client.

DATA FLOW DIAGRAM

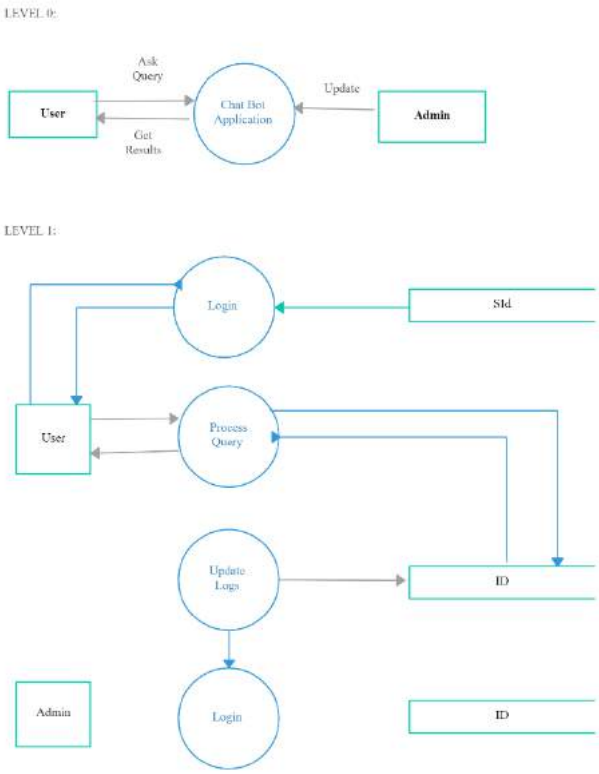


Fig:3 - Data Flow Diagram

IV. Experimental Evaluation

Based on the databases, python dependencies need to be installed. Its configuration is defined in endpoints. The data needed to be fetched from the database, this processing can be done using set of defined actions. An instance of connection on which cursor is defined, is used to execute the query. Based on the result from database, the response is given to end user.[1]

From the study it can be seen that rasa core features like slots, forms, supervised interactive learning, API integration, and database makes it a complete framework that can be used to perform highly complex tasks. The chatbot based on rasa has more capabilities than any open-source alternative.[2]

Further, in this paper, internals of rasa have been modified to carry out custom data logging of client IP

and port. All internals and custom action have been studied which further states that rasa is a complete open-source framework for the development of chatbots and for the developers who don't want to dig into the internals of natural language processing. The future scope of this study, voice and face recognition engines can be integrated for more complex task like ATM cash withdrawal. Performance may be enhanced with use of various learning procedures of machine learning.

Result:

To analyse users queries and understand users' message. To provide an answer to the query of the user very effectively. It also helps to save the time of the user since s/he does not have to personally go to the college for inquiry.

This system will help the student to be updated about the college activities. The system will reply using an effective GUI which implies that as if a real person is talking to the user. College admission related queries could be answered through it. Viewing user profiles and retrieves attendance and grade/ pointers.

College students can get information about examinations to be held. College students can fetch particulars about placement activities.

V. CONCLUSION

The goal of the system is to help the students to stay updated with their college activities. Artificial Intelligent is the fastest growing technology everywhere in the world, with the help of Artificial Intelligent and Knowledgeable database. We can make the transformation in the pattern matching and virtual assistance. This system is developing chat bot based on android system so with the combination of Artificial Intelligent Knowledgeable database and virtual assistance. We can develop such chat bot which will make a conversion between human and machine and will satisfy the question raised by user. The main motive of the project is to reduce the work load on the college's office staff and reduce the response time to a user's query.

VI. REFERENCES

[1]. Jayesh Gangrade1, Surinder Singh Surme2, Sumant Somu3, Shubham Raskonda4, Poonam Gupta, A Review on College Enquiry Chatbot, G.H. Raisoni College of Engineering and Management, Pune, India, 2019 IJESC

[2]. Hiral Paghadal1, Anezka Virani2, Apratim Shukla3, Dr. G T Thampi, Department of Information Technology, Thadomal Shahani Engineering College, Mumbai, Maharashtra, India, May 2020 www.irjet.net

[3]. Ms.Ch.Lavanya Susanna, R.Pratyusha,P.Swathi, P.Rishi Krishna,V.Sai Pradeep, "College Enquiry Chatbot", InternationalResearch Journal of Engineering an Technology (IRJET), 3 Mar2020

[4]. Assistant Prof Ram Manoj Sharma, "Chatbot based College Information System", RESEARCH REVIEW International Journal of Multidisciplinary, Issue- 03, March-2019

[5]. P.Nikhila, G.Jyothi, K.Mounika, Mr. C Kishor Kumar Reddy and Dr. B V Ramana Murthy on , "Chatbots Using Artificial Intelligence", International Journal of Research and Development, January/2019

[6]. Payal Jain, "College Enquiry ChatBot Using Iterative Model", International Journal of Scientific Engineering and Research (IJSER),ISSN (Online):January 2019,

[7]. Sagar Pawar, Omkar Rane, Ojas Wankhade, Pradnya Mehta, "A Web Based College Enquiry Chatbot with Results", International Journal of Innovative Research in Science, Engineering and Technology, April 2018

[8]. Jincy Susan Thomas, Seena Thomas, "Chatbot Using Gated End-to- End Memory Networks", International Research Journal of Engineering and Technology (IRJET) 03 Mar 2018

[9]. Harsh Pawar , Pranav Prabhu, Ajay Yadav, Vincent Mendonca , Joyce Lemos, "College Enquiry Chatbot Using Knowledge in Database", International Journal for Research in Applied Science & Engineering Technology (IJRASET), April 2018,

[10]. Prof. Suprita Das, Prof. Ela Kumar, "Determining Accuracy of Chatbot by applying Algorithm Design and Defined process", 4 th International Conference on Computing Communication and Automation (ICCCA), IEEE,2018

[11]. Prof.K.Bala, Mukesh Kumar ,Sayali Hulawale, Sahil Pandita,"Chatbot For College Management System Using A.I", International Research Journal of Engineering and Technology (IRJET) Nov - 2017

[12]. Nitesh Thakur, Akshay Hiwrale, Sourabh Selote, Abhijeet Shinde and Prof. Namrata Mahakalkar, "Artificially Intelligent Chatbot", Universal Research Reports, July - September 2017.

[13]. Jayesh Gangrade, Surinder Singh Surme, Sumant Somu, Shubham Raskonda, Poonam Gupta on "A Review on College Enquiry Chatbot" , International Research Journal of Engineering and Technology (IJESC), Issued Nov 2019.

[14]. Colace, F., De Santo, M., Lombardi, M., Pascale, L., Pietrosanto, A. (2018). Chatbot for E-Learning: A Cases Study. International Journal of Mechanical Engineering and Robotics Research Vol. 7, No. 5, September.

[15]. Ashok, G., Brian, C., Mithun, K., Shanu, S., Abhinaya, S., & Bryan, W. (2015). Using Watson for Enhancing Human-Computer Co Creativity. AAAI Symposium: 22–29.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Patient Feedback using Speech Emotion Recognition

Rutuja Patil[1], Siddhi Salunke[1], Pournima Ubale[1], Mayur Talole[1], Prof. Ajita Mahapadi[2]

[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

[2]Assistant Professor, Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

Hospitals attempt to collect feedback from the patients to study their sentiment towards their services and facilitates provided by the hospitals to improve their environment. In present scenario feedbacks are taken in written form and are not truly maintained by hospital staff , and this technique does not reveal the true sentiments of the patients ,but this SER feedback provides a chance to highlight certain aspects. In this paper, a method has been proposed for emotion recognition by speech based on speech features and speech transcriptions, such as Spectrogram, Mel-Frequency Cepstral Coefficients (MFCCs) ,helps to retain emotion related low level characteristics in speech where as text helps capture semantic meaning both of which help in different aspects of emotion detection.

Keywords: Speech Emotion Recognition, SER, Speech Transcriptions, Speech Features .

I. INTRODUCTION

Feedback is an event that occurs when the output of a system is used as input back into the system as part of a chain of cause and effect. Feedback plays an important role in hospital by helping to adopt new knowledge and prevent repetitive mistakes. Feedback is a process which helps the organization to monitor, evaluate, and regulate the overall working environment. Good feedback practice provides useful information to the organization in improving the teaching and learning experience.

Traditionally feedback in hospital is filled out manually through forms due to which patients pay no attention about filling the form seriously. Also this process is time-consuming and very tedious job. Their might be also duplications of data and the information entered maybe false or misleading . There are many chances to lose data. Hospital might try to maintain only the positive feedback for its reputation , and this might be misleading to the people. This paper focuses on generating Patient feedback that takes voice as an input, analyze it using Speech Emotion Recognition(SER) & conveys us the feedback of the patients through their emotions. Also it generates ratings through feedback such as Excellent ,satisfactory , non- satisfactory, need improvement etc.

© 2020 IJSRST | Print ISSN: 2395-6011 | Online ISSN: 2395-602X | Published : 18/12/2020 - 216-220

Emotion Recognition using speech(SER)

- Speech is the most important feature communication tool .
- It is a biometric feature like fingerprint and carries the emotional state of the speaker.
- Therefore, speech data extracted from real talks may result more realistic emotional features than textual data.

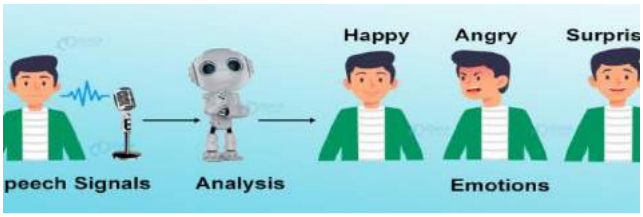


Fig1. Speech Emotion Recognition

- Speech Emotion Recognition (SER) systems is a collection of methodologies in which speech signals are classify and process to detect the emotions.

- <u>Applications of SER</u>
  - Human communication with machine
  - Psychological consultation
  - Patient care
  - Call centers

- <u>Objective:</u>
  - To automate the work of taking the feedbacks of patients.
  - To simplify the task of Patients.
  - To efficiently organize the feedback record.
  - To reduce errors and duplication of records.
  - To maintain True data

II. LITERATURE SURVEY

Sr. No	*Paper*	*Remarks*
1.	Review on Emotion Recognition using Speech	features of speech can be used to identify the difference between several emotional statements
2.	Speech Emotion Recognition using deep learning.	Some of the classifications algorithms like K-NN, Random Forest are used to classify Emotions
3.	Deep Learning Based Emotion Recognition system using Speech features and Transcriptions	Transcription and Speech Features such as Spectrogram and MFCC help to retain the Emotion.

III. PROPOSED METHOD

This paper focuses on speech features and speech transcriptions such as Spectrogram and MFCC, which together provide low-level features required and the necessary semantic relationships to classify among different emotions correctly.

We perform experiments on both speech features and speech transcriptions individually and also together, which gives us greater accuracies than the previous methods. Various combinations of data has been used as inputs in different Deep Neural Network(DNN) architectures, the details of architectures is show below.

1. CNN model based on transcriptions (Model 1)

Transcription is a printed or written version of something. Speech transcriptions is transcribes a spoken audio into text and returns block of text with its semantic meaning for each portion of the transcribed audio.

Speech transcriptions plays an important role in recognising emotions using speech. It is necessary because because it is hard to know the context in which this word has been used.

For e.g.

"good"-can have positive meaning.

-can be used in sarcastic way.

The CNN model described in this section takes speech transcriptions, in the form of word embeddings, as input to detect emotion. CNNs can directly be applied to word embeddings without prior information on their semantic contexts. This strongly suggests that language can be interpreted as a signal no different from other signals.

A word embedding is trained representation for text where words that have similar meaning have a same representation.They are distributed representation of text that is one of the key breakthroughs for the impressive performance of deep learning methods on challenging natural language processing(NLP) problems.

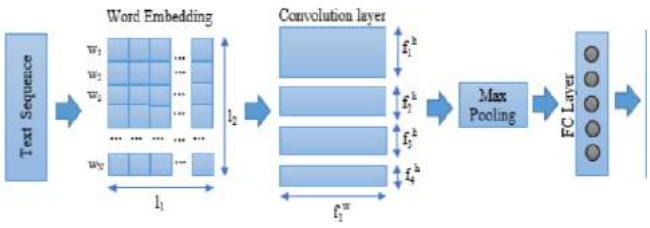


Fig 2. Transcriptions–based on CNN model

The architecture of the Transcription-based CNN model as shown in the fig 2 is used in this paper for our purpose.

Transcription sequences (embedded vectors), which given as an input to this model, are convolved with kernels of different sizes. The utrmost number of the words in any given utterance has been set to 128, which covers entirety of the IEMOCAP dataset. One feature from each of the different convolutional layers is picked by the max-pool layer. These features are fed to the single FC layer of our model. Finally, a softmax layer is used to perform classification. We experimented with batch-normalization, a technique which helps prevent the model from over-fitting and also from being too sensitive to the initial weight, and also varied dropout rates from 0.25 to 0.75. An improvement in convergence rate is also observed with the use of batch-normalization.

2. CNN model based on speech features (Model 2)

Spectrograms and MFCCs are commonly used to represent speech features.

Spectrogram:

A spectrogram displays signal strength over time at the various frequencies present in a waveform. Spectrograms can be two-dimensional graphs with a third variable represented by color, or three-dimensional graphs with a fourth color variable.

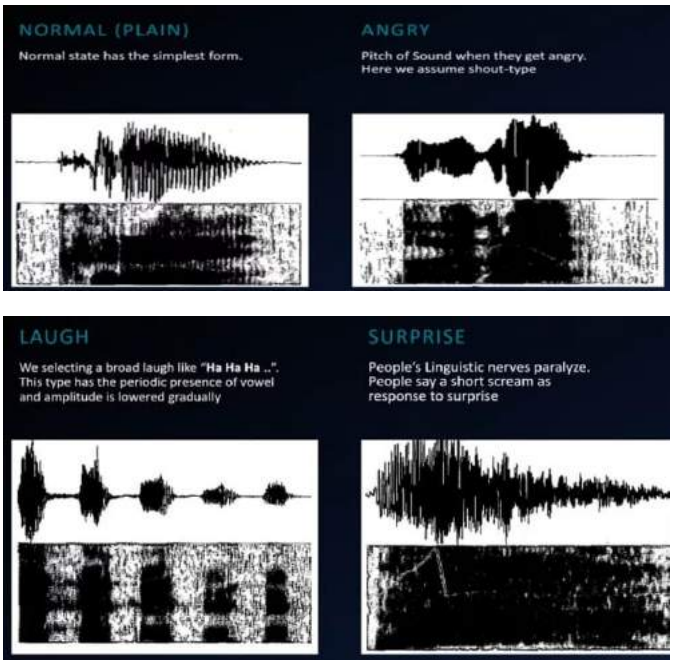


Fig 3. Spectrograms

Fig 3 represents different spectrogram for the different pitch or for different Emotions.

The obtained Spectrogram magnitudes are then mapped to the Mel-scale to get Mel-spectrograms. 128 Spectrogram coefficients per window are used in this model. The Mel-frequency scale puts emphasis on the lower end of the frequency spectrum over the ones which are higher , which imitates the perceptual hearing capabilities of the humans. Along with the parameters mentioned above , we have used "librosa" python package to compute the mel-spectrograms.

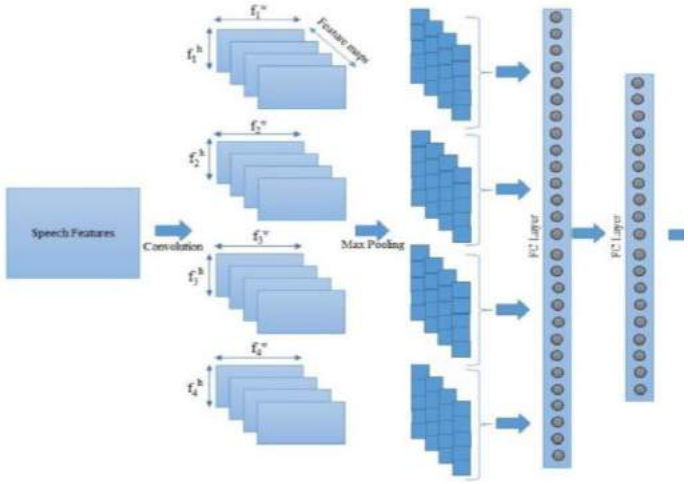


Fig 4 Spectrogram/MFCC based CNN model

Fig. 4 explains the 2D CNN architecture which is used to detect emotion using Spectrograms . A set of 4 parallel-2D convolutions are applied on the Spectrogram to extract speech features. The input shape of the Spectrogram image is 128 x 256 (number of Mel's x number of windows). 200 2D-kernels are used for each of the parallel convolution steps. Figuring out the optimal kernel size is a difficult and time taking task, which may depend on several factors all which cannot be clearly defined. To prevent choosing one single kernel size that could possibly be sub-optimal we decided to use kernels of different sizes, each of which is fixed for a single parallel path, so that we can take advantage of the different patterns that are chosen by each of the kernel. The sizes of each of the kernels in their respective parallel CNN paths are 12 x 16, 18 x 24, 24 x 32, and 30 x 40. The features generated in the said convolution layers are then fed to their respective max-pool layers, which extracts 4 features from each filter as the pool size is exactly half along the width and height of the convolution output. The extracted features are fed to the Fully Connected (FC) layer. This model makes use of two FC layers of sizes 400 and 200. Batch normalization is applied to both the FC layers. We experimented with dropout rates varying between 25% and 75% for the first FC layer but excluded it completely from the second FC layer.

3. Combined CNN model based on both speech transcriptions and speech features(model3)

A text-based CNN model fails to capture the low-level features of speech signals and due to which it cannot achieve a very high emotion detection accuracy. As mentioned above the combined Spectrogram and MFCC based CNN model achieves an improvement of 4% over existing state-of-the-art methods. The combined Text-MFCC model performs even better and beats the benchmark class.

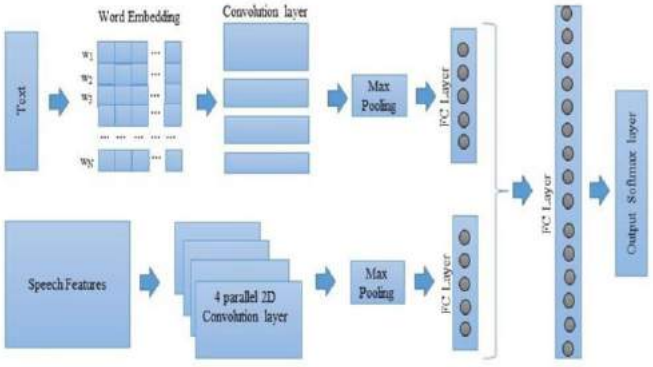


Fig5. Combining Speech features and spectrogram

In Model 3 the Spectrogram channel consists of 4 parallel 2D-CNN layers with kernels of different sizes. MFCC channel also consists of 4 parallel 2D-CNN layers as compared to the Spectrogram channel. Outputs from Spectrogram and MFCC channels are fed to one FC layer each. The outputs of both the FC

layers, after normalization, are concatenated and fed to the 2nd FC layer. The final step is to feed the outputs of the last FC layer to a softmax layer.

IV. ARCHITETURE DIAGRAM

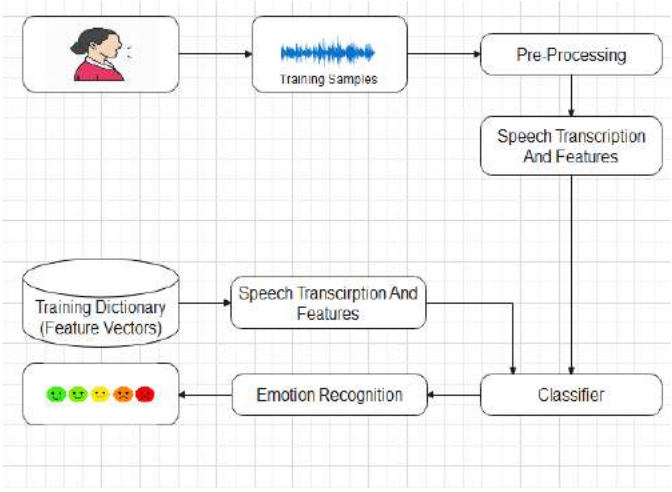


Fig 6. Architectural Diagram

V. DATASET

We are using IEMOCAP dataset [14] for proposed method. The IEMOCAP dataset consists of five sessions from both scripted and spontaneous act. From total 9 categories of emotion, only 3 emotion categories are used i.e. anger, excitement, neutral and sadness. 4936 out of 10039 turns is the total number of utterance used. The distribution of utterances for each class is almost identical to make the dataset balance. From many modalities, only speech and text are used on this research. Speech signals which are present in the dataset are processed at sampling rate of 16kHz with average length 4.5s. and for the text, average words per turn is 11.4 words while longest utterance has 554 words. Text and speech not aligned when it is processed for feature extraction. The processing for both modalities is performed independently and simultaneously through its networks.

VI. CONCLUSION

Therefore, in this paper multiple CNN based architecture have been proposed which includes speech features and speech transcriptions . model 1 provides greater accuracy , which then later improves greater accuracy when it is combined with model 2 .The model 3 which include both speech transcriptions(model 1) and speech features(model 2) results in an overall emotion detection accuracy greater than previous model by 7% , that is 76.10%. The proposed model can be used to detect emotions of the patients based on their feedback using speech.

VII. REFERENCES

[1]. Suraj Tripathi1, Abhay Kumar1*, Abhiram Ramesh1*, Chirag Singh1*, Promod Yenigalla1, “Deep Learning based Emotion Recognition System Using Speech Features and Transcriptions” , Samsung R&D Institute India – Bangalore , arXiv.org, 2019.

[2]. Nithya Roopa S., Prabhakaran M, Betty.P, Nov 2018. “Speech Emotion Recognition using Deep Learning”. International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7 Issue-4S, November 2018

[3]. k Ashok Kumar, J L Mazher Iqbal . “Machine Learning Based Emotion Recognition using Speech Signal” . International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-1S5, December, 2019

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Automatic Answer Sheet Checker

Pratik Trimbake[1], Swapnali Kamble[1], Rakshanda Kapoor[1], Vishal Kisan Borate[2]

[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

[2]Assistant Professor Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

An automating the task of scoring subjective answer is considered. The goal is to assign score which are comparable to those of human score by coupling AI technologies . In this process involves many image level operation i.e. removal of pre-printed matter , extraction and segmentation of words. Scoring is based on machine learning of parameter and natural language processing. System checks answer and score as good as human being. We present an Answer Sheet Checker based on Textual Entailment and Question Answering. The important features used to develop the Answer Sheet Checker System are named Entity Recognition, Textual Entailment, Question-Answer type Analysis and Chunk Boundary and Dependency relations. Separate Answer Sheet Checker modules have been developed for each of these features. We first combine the question and the supporting text to check the entailment relations as either "VALIDATED" or "REJECTED". Once the user enters his/her answers the system then compares this answer to original answer written in database and allocates marks accordingly. The system requires you to store the original answer for the system. This facility is provided by the Admin. The admin may insert questions and respective subjective answers in the system. When a user takes the test he/she is provided with questions and areas to type his answers. The purpose of this system is to automate the old fashioned manual system and introduce automatic evaluation of marks in much faster and accurate way.

Keywords - Data-mining, Stop word Selection, Text Classification, Stemming Algorithm and Stripping Algorithm.

I. INTRODUCTION

The answer sheet is widely used for student performance in exam in school and college .The main approach is to evaluation is efficient and reliable. An automatic answer sheet checker checks the answer sheet and written mark as similar to human being .This software is built to check the subjective answer. The system consist of in build artificial sensor that verify answer and allocate marks according as good as human being accessing large number of handwritten answer sheet is relatively time consuming task there is an intense need of speed up and enhance a process of rating handwritten words while maintaining cost effectiveness .It is relatively inexpensive answer written by hand .The primary means of testing the student on state assessment of reading compression motivation of these system is

mainly always we have seen the online OMR sheet checker or objective answer sheet checker but the main goal is to develop subjective answer checker. Artificial Intelligence is an ability to design smart machines or to develop self- learning software applications that imitate the traits of the human mind like reasoning, problem-solving, planning, optimal decision making, sensory perceptions etc.

Contribution:

In this paper, we proposed the method in which the Examiners get bored by checking many answer sheets, hence the system reduces their workload by automating the manual checking process accurately. The system calculates the score and provides results instantly. It removes human errors that commonly occur during manual checking. The purpose of this system is to automate the old fashioned manual system and introduce automatic evaluation of marks in much faster and accurate way.

Structure:

This paper is organized as follows: Section 2 presents Related Work. Section 3 summarizes our Proposed Work which consists of our base algorithm which presented in [1],[2],[3]. We will evaluate our methods and compare with existing work in Section 4 as Experimental Setup and Results. Finally, in Section 5 we conclude and present suggestion for future work.

II. RELATED WORK

1] "Case Based Modeling of Answer Points to Expedite Semi-Automated Evaluation of Subjective Papers" paper is proposed by Chhanda Roy , Chitrita Chaudhuri. In this paper the primary goal of this paper is to propose a framework, where textual papers set for subjective questions, are supplemented with model answer points to facilitate the evaluation procedure in a semi-automated manner. The proposed framework also accommodates provisions for reward and penalty schemes. In the reward scheme, additional valid points provided by the examinees would earn them bonus marks as rewards.

2] "Automatic answer sheet checker" paper is proposed by Gunjal M.S , Sanap K.N , Sable R.G , Nannaware P.S, Ghuge R.B. In the paper an automating the task of scoring subjective answer is considered. The goal is to assign score which are comparable to those of human score by coupling AI technologies .In this process involves many image level operation i.e. removal of pre printed matter , extraction and segmentation of words. Scoring is based on machine learning of parameter and natural language processing. System checks answer and score as good as human being.

3] "Automatic Answer Validation System on English Language" paper is proposed by Partha Pakray, Santanu Pal, Sivaji Bandyopadhyay. In this paper an Answer Validation System (A V) based on Textual Entailment and Question Answering. The important features used to develop the A V system are Named Entity Recognition, Textual Entailment, Question-Answer type Analysis and Chunk Boundary and Dependency relations. Separate A V modules have been developed for each of these features. We first combine the question and the answer into Hypothesis (H) and the Supporting Text as Text (T) to check the entailment relation as either "VALIDA TED" or "REJECTED". The important lexical features used for the Textual Entailment Module in the present system are: WordNet based unigram match, bigram match and skip-gram. The results obtained from the four answer validation modules are integrated using a voting technique. Evaluation scores obtained on the AVE 2008 test set show 67% precision for "VALIDATED" decisions.

© 2020 IJSRST | Print ISSN: 2395-6011 | Online ISSN: 2395-602X | Published : 18/12/2020 - 221-226

III. PROPOSED WORK

Working:

Online examination is the use of the Internet for examination activities, or we can say, it is the process by which a student learns via the Internet with the help of a subject expert or an examination. How it works? This examination technique involves use of several web- based programs to achieve its goals. Some of these programs or applications are email, an instant messaging, online whiteboards, etc. An automatic answer checker application that checks and marks written answers similar to a human being. This software application is built to check subjective answers in an online examination and allocate marks to the user after verifying the answer. The system requires you to store the original answer for the system. This facility is provided to the admin. The admin may insert questions and respective subjective answers in the system. These answers are stored as notepad files. When a user takes the test he is provided with questions and area to type his answers. Once the user enters his/her answers the system then compares this answer to original answer written in database and allocates marks accordingly. Both the answers need not be exactly same, word to word. The system consists of in built artificial intelligence sensors that verify answers and allocate marks accordingly as good as a human being. The system will consist of the following elements: Login: The proposed system will have two login facility: User login: The user login is the login allocated for the students. As soon as you click the student login button you will be asked to enter login id and password. The system will check for the id and automatically display students name, email id and phone num for verification. The user login will be able to write answers with respect to the question uploaded. The system will show marks scored as soon as you enter the next button Admin login: The admin login will let the teacher's login. The admin login's each user will have his own password and id through which they can login in. The admin can add subtract questions, check for students marks and so on. Just like the teachers can do manually. Answer checking: Suppose the question is "who is hitler?" Case 1: Was a cruel ruler found the narzi party. In this case the anwer will get 10 points for correct answer Case 2: Was a cruel ruler. Found the BJP. In this case the statement 1 is right and 2 is wrong so will be given 5 points out of 10 as only half of the answer is correct. Thus depending on answers the points of the answers will be set by the admin.

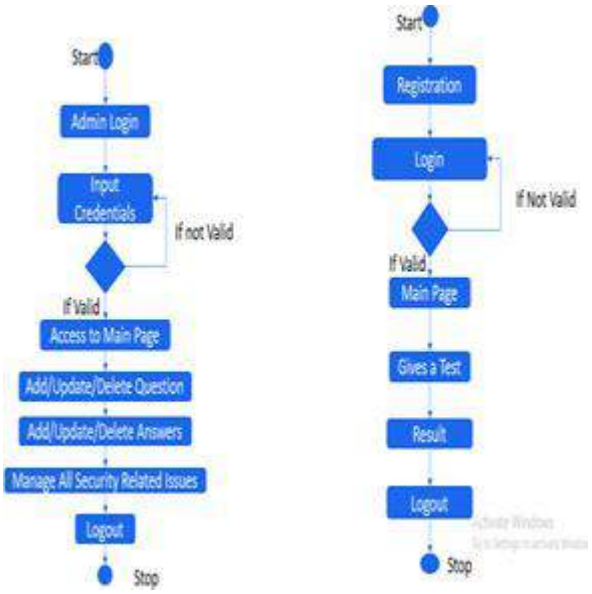


Fig 1. Activity Diagram for Admin and Student.

Proposed Algorithm:

1] Key word search algorithm: A search algorithm is an algorithm that retrieves information stored within some data structure. Data structure can include linked list, array, search tree, hash table or various other storage methods the appropriate search algorithm often depends on the data structure being searched. Searching also encompasses algorithm that query the data structure such as SQL SELECT command. Search algorithm can be classified based on their mechanism of searching. Linear search algorithm

check every record for the one associated with target key in a linear fashion. Binary search repeat target the center of the search structure and divide the search & digital search algorithm. Hashing directly maps keys to record based on a hash function. searches outside of a linear search require that the data be sorted in some way. Search functions are also evaluated on the basis of their complexity or maximum theoretical runtime. Keyword search: Keyword search is the most popular information discovery method because the user does not need to know either a query language or the underlying structure of the data. The search engine are available today provide keyword search on top of sets of document when a set of keyword is provided by the user search engine return all document that are associated with these keywords. Typically two keyword & a documents are of associated with keywords are contained in the document & their degree of associatively is often distance from each other. Keyword research is a practices search engine optimization professionals use to find & research actual search terms that people enter into search engine optimization professional research keyword which they use to achieve better ranking in search engines.

2] Stemming Algorithm: It is the process for removing the commoner morphological & in flexional ending from words in English. It is main use is as part of a term normalization process that is usually done when setting up information retrieval system. Stemming refers to the process of removing affixes (prefixes & suffixes) from words. In the information retrieval context, stemming is used to conflate word from to avoid mismatches that may undermine recall. As a simple example consider searching for a document entitled "How to write" if the user issues the query "writing " there will be no match with the title .however if the query is stemmed so that "writing "becomes "write" then retrieval will be successful. Stemming is the process of finding the route word.

Given below is the System Architecture of this Automatic Answer Sheet Checker:

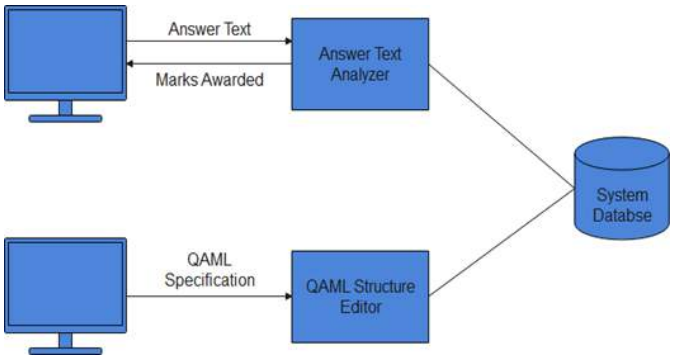


Fig 2. System Architecture

DATA FLOW DIAGRAM

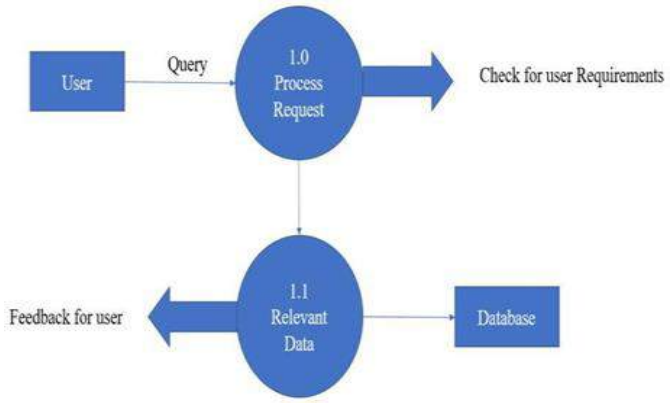


Fig 3. DFD Level 1

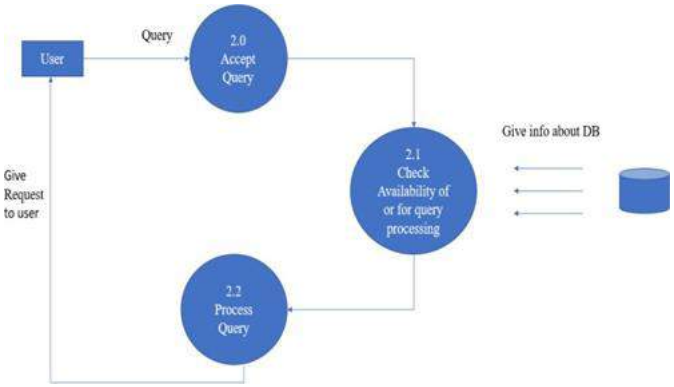


Fig 4. DFD Level 2

IV. Experimental Evaluation

Based on the databases, Artificial Intelligence needs to be installed. Incorporating AI into complex systems poses numerous challenges to traditional test and evaluation methods. As AI handles varying decision levels, it needs conference to ensure testable, repeatable and auditable decisions. Additionally, we need to understand failure modes and failure mitigation techniques. The key challenges embedded-AI exacerbates, the themes based for how traditional and evaluation will evolve to provide AI system assurance.

Result:

The current manual evaluation takes about 60 seconds to evaluate an answer whereas the proposed system takes about 15 seconds to evaluate an answer. The proposed system is 300% more time efficient as compared to manual answer evaluation system. The proposed system is about 75 - 87.5% accurate with comparison to manual system. The proposed system completely eliminates the human effort and time to evaluate an answer.

Proposed system can evaluate 5760 answers in a day where as a human working for 8 hours can evaluate 480 answers a day . Hence, proposed system can evaluate 1100% more answers compared to that of manual evaluation system.

V. CONCLUSION

In this paper an attempt has been taken to present an analysis and survey on AUTOMATIC ANSWER CHECKER and important protocols proposed by researchers with their Answer sheet checking. This System calculates the score and provides results instantly for descriptive answers. This System is design to reduce the burden of checking the papers manually.

VI. REFERENCES

[1]. Partha Pakray, Santanu Pal and Sivaji Bandyopadhyay “Automatic Answer Validation System on English Language” IEEE 2010.

[2]. Gunjal M.S., Sanap K.N., Sable R.G., Nannaware P.S., Ghuge R.B. , “Automatic answer sheet checker. ”, International Journal of Advanced Engineering and Science research(IJAES) ,Volume 5, Issue 1, March 2017.

[3]. Chhanda Roy , Chitrita Chaudhuri “Case Based Modeling of Answer Points to Expedite Semi-Automated Evaluation of Subjective Papers” IEEE 2018.

[4]. http://eduexamsoftware.weebly.com.

[5]. www.projectcorner.in/online-examination-system- college -project-asp-net.

[6]. http://oes.sourceforge.net.

[7]. www.codeguru.com

[8]. Web Enabled Commercial Appplication Developmen -Ivan Bayross .

[9]. Head First Servlets and J SP -Bryan Basham,Kathy Sierra and Bert Bates.

[10]. V. Senthil Kumaran and A Sankar, “Towards an automated system for short-answer assessment using ontology mapping”, International Arab Journal of e- Technology, Vol. 4, No. 1, pp: 17-24, January 2015

[11]. H. Mittal and and M. Syamala Devi, “Computerized Evaluation of Subjective Answers Using Hybrid Technique”, International Conference on Innovations in Computer Science and Engineering (ICICSE 2015), Part of the Advances in Intelligent Systems and Computing book series, volume 413, pp: 295-303,

[12]. R. C. Schank, ”Dynamic Memory : A theory of reminding and learning in computers and people”,

Cambridge, UK: Cambridge University Press, 1982

[13]. R. C. Schank, "Memory-based expert systems", Technical Report (# AFOSR. TR. 84-0814), Yale University , New Haven, USA, 1984

[14]. K. D. Ashley, "Case-based reasoning and its implications for legal expert Systems", Artificial Intelligence and Law, Volume 1, Issue 2-3, pp 113-208, June 1992

[15]. M. Nilsson, M. SollenBorn, "Advancement and Development Trends in Medical Case-Based Reasoning : An Overview of Systems and System Development", © American Association for Artificial Intelligence, 2004

[16]. D. W. Aha, "The omnipresence of case-based reasoning in science and application", Knowledge- Based Systems, vol. 11, no. 5-6, pp. 261-273, 1998

[17]. T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, "Introduction to Algorithms", 2nd Edition, Prentice Hall of India (Eastern Economy Edition), 2001

[18]. T. Nawaz, K. I. Qazi, and MdI. Ashraf, "An Efficient Algorithm of Computerized Checking System for Hard Copy MCQs Test(HCMCQST)", IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.5, pp: 228-236, May 2009

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Bank Locker Security System using Machine Learning with Face and Liveness Detection

Akash Mote[1], Kanhaiya Patil[1], Akshay Chavan[1], Mrunal Saraf[1], Prof. Ashwini Pandagale[2]
[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India
[2]Professor, Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

Ensuring the security of transactions is currently the biggest threat facing banking systems. The use of biometric authentication of users attracts huge sums of money from banks around the world due to their convenience and acceptance. Especially in offline environments, where face images from ID documents are matched to digital selfies. In fact, comparisons of selfies with IDs have also been used in some broader programs these days, such as automatic immigration control. The great difficulty of such a process lies in limiting the differences between comparative facial images given their different origins. Based on deep features extracted by two well-referenced Convolutional Neural Networks(CNN).we suggest a novel architecture for cross-domain matching problem . The results obtained from the data collected, called Face Bank, with more than 93% accuracy, indicate the strength of the proposed face-to-face comparison problem and its inclusion in real banking security systems.

Keywords: Convolutional Neural Networks(CNN), Face Bank, automatic immigration control, Digital selfies, Face-to-face comparison problem.

I. INTRODUCTION

Much work is still necessary to allow convenient, secure and friendly systems to be designed.the recognition performance of biometric system is satisfactory for most applications, In face recognition, the usual attack methods may be classified into certain categories.

The concept of classifying is placed on what verification proof is equip to face verification system, such as a lifted photo, lifted face photos, recorded video, 3D face models with the skills of blinking and lip moving, 3D face models with various expressions and so on. The concept of classifying is placed on what verification proof is provide to face verification system, like a stolen photo, stolen face photos, recorded video, 3D face models with the skills of blinking and lip moving, 3D face models with

various expressions and so on. In this paper, we proposed a method of live face detection to resist the attack using a photograph.on what verification proof is provide to face verification system, such as a lifted photo, lifted face photos, recorded video, 3D face

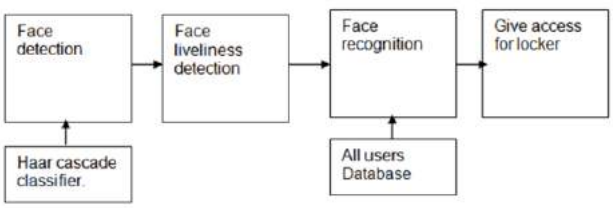
models with the skills of blinking and lip moving, 3D face models with various expressions and so on. The concept of classifying is placed on what verification proof is provide to face verification system, such as a stolen photo, lifted face photos, recorded video, 3D face models with the skills of blinking and lip moving, 3D face models with various expressions and so on. In this paper, to resist the attack using a photograph we proposed a method of live face detection Our algorithm is placed on analysis of movement of facial components, especially eyes, in sequential images.In sequential face images there are very small variations in shape of face and facial elements. But eyes have larger variation in shape because we always blink,move the pupils unconsciously. So we spot eyes in sequential face images and compare the shape of each eye region to decide whether the input face image is a real face or a photograph.

A. Problem Statement

With the demand of face recognition, criminals will try to attack the face recognition system, for which liveness detection has become an crucial part of the authentication system. Among the current liveness detection algorithms, methods based on machine learning . Therefore, we proposed this method in this paper.

B. Model Framework

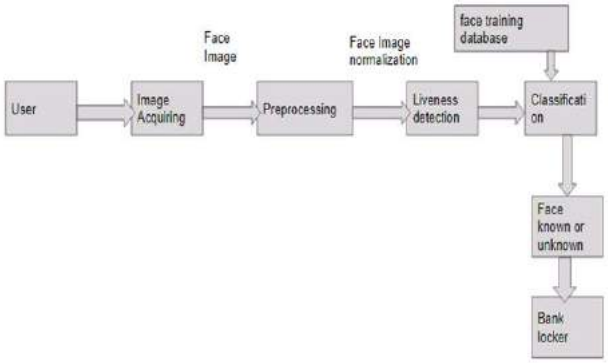
The proposed framework that combines Face Net with liveness detection is shown in the figure.



In above block diagram we are going to detect face using haar cascade classifier which algorithm for detection of face. After detection of face, system will decide the face is real or fake by using liveness detection technique. Liveness detection technique is the act of differentiating the feature space into live and non-living

In this system we need a way to detect faces and eyes in real-time. So we are using -cascade classifier to performs these tasks. In this haar cascade classifier Cascade is a machine learning object detection algorithm used to identify objects in an image or video.

C. Architecture Diagram



In this diagram we are going to implement eye-blink detection & face recognition Based on LBPH algorithm. The algorithm works in real time through a webcam and displays the human's name. The program runs as follows:

1. Detect faces in each frame generated by the webcam.
2. For each detected face, detect eyes.
3. Detect liveness of the face i.e. eyes are blinking or not
4. Recognize face and access the respected locker of the user.

II. LITERATURE REVIEW

Gang Pan et al.[1] present a spoofing against photograph in face recognition using real time liveness detection using spontaneous eye blinking. This method requires only a generic camera no other hardware to avoid spoofing attack in nonintrusive manner. Eye blinking is physical process which instantaneously opens and closes lids many times in a minute. Generic camera captures 15 frames per seconds, it gives two frames of faces which used as clue against spoofing attack. Two captured frames in multiple frames to check liveness, so user should be co-operative.

Face liveness detection [3] has been proposed to improve the reliability and security of face recognition system. The fake faces are distinguished from the real ones using different classification techniques.

In this paper, we propose a single image-based fake face detection method based on frequency and texture analyses for discriminating 2-D paper masks from the live faces. For the frequency analysis, we have carried out power spectrum based method [4] which exploits not only the low frequency information but also the information residing in the high frequency regions. Moreover, widely used Local Binary Pattern (LBP) [5]. In face recognition, the usual attack methods may be classified into various different types. The basic ideology of classifying is based on what verification proof is provided to face verification system, such as a stolen face photos, recorded video, 3D face models, fake photos with the abilities of blinking and lip moving, 3D face models with various expressions and many more to mention[6].

The main goal of our paper is to design and implement a bank locker security system solely based on RFID and GSM technology which has to be organized in the bank, secured offices and homes. In our system only authentic person can be the one recovering the money from the bank locker. The sequence are considered as independent. HMM produces features from finite state set. Typical eye blink activity using HMM feature finds spoofing attack.

Anjos et al. [2] proposed a method based on foreground or background motion correlation for checking liveness of user. This method classified in motion detection. This method works on correlation between head rotation of user and its background. To find correlation author uses fine grained motion direction. Optical flow is used to find the direction of motion. This approach is easy process but require RFID reader reads the identification number from passive tag and sends it to the microcontroller, if the identification number is valid then microcontroller sends the SMS request to the authenticated person mobile number, for the original password to open the bank locker, if the person enters the password to the microcontroller which will verify the passwords entered by the key board and received from authenticated mobile number . If these two passwords are matched the locker will be opened otherwise it will be remain in locked position[7].

Initially pattern flow are collected as the datasets are maintained in bank agent server. The device has a camera to capture the pattern flow of user and sends it for processing, features of the logic were compared and user was recognized. In addition to the authentication of user there is another system to identify the user before that RFID tag checking is needed. Image processing is used and the keypad password is needed for another level of security. In future bank can implement this type of authentication option for banking and from this project it shows that all the bank accounts can be accessed without using cards through this face recognition efficiently and safely [8].

Access control system forms a essential link in a security chain. The biometric password and pattern based security system presented here is an access control system that allows only the authorized persons to access a restricted area. We have implemented a locker security system based on fingerprint, password and GSM technology

containing door locking system which can activate, authenticate and validate the user and unlock the door in real time for locker secure[9].

They say perhaps the most important application of accurate personal identification is acquiring limited access systems from malicious attacks. Among all the presently employed biometric techniques, biometric identification systems have received the most attention due to the long history of fingerprints and their extensive use in forensics. Our paper deals with the issue of selection of an optimal algorithm for biometric suiting in order to design a system that matches required specifications in performance and accuracy[10].

III. CONCLUSION

In this paper, we have proposed a machine learning based face detection-recognition and liveness detection for bank locker. It is highly reliable system to ensure the security of our valuables.

IV. REFERENCES

[1]. G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink - based anti-spoofing in face recognition from a generic webcamera," in Proc. IEEE 11th Int. Conf. Comput. Vis. (ICCV), Oct. 2007, pp. 1–8.

[2]. Anjos, M. M. Chakka, and S. Marcel, "Motion-based countermeasures to photo attacks in face recognition," IET Biometrics, vol. 3, no. 3, pp. 147–158, Sep. 2014.

[3]. Pan, Gang, Lin Sun, Zhaohui Wu, and Yueming Wang. "Monocular camera-based face liveness detection by combining eyeblink and scene context." Telecommunication Systems 47, no. 3-4 (2011): 215-225.

[4]. H. S. Choi, R. C. Kang, K.T. Choi, A. T. B. Jin, and J.H. Kim. Fake-Fingerprint Detection using Multiple Static Features. Optical Engineering, 48(4), 2009.

[5]. T. Ojala, and M. Pietikainen. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24

[6]. J. Li, Y. Wang, T. Tan, and A. K. Jain, "Live face detection based on the analysis of fourier spectra," In Biometric Technology for HumanIdentification, SPIE vol. 5404, pp. 296-303, 2004.

[7]. Z. Lu, X. Wu, and R. He, "Person identification from lip texture analysis," in International Conference on Digital Signal Processing, DSP, 2017, pp. 472–476.

[8]. Gan, J.Y.; Li, S.L.; Zhai, Y.K.; Liu, C.Y. 3D convolutional neural network based on face anti-spoofing. In Proceedings of the International Conference on Multimedia and Image Processing, Wuhan, China, 17–19 March 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–5.

[9]. Li, L.; Feng, X.Y.; Jiang, X.Y.; Xia, Z.Q.; Hadid, A. Face antispoofing via deep local binary patterns. In Proceedings of the IEEE International Conference on Image Processing, Beijing, China, 17–20 September 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 101–105.

[10]. Wang, S.Y.; Yang, S.H.; Chen Y, P.; Huang, J.W. Face liveness detection based on skin blood flow analysis. Symmetry 2017, 9, 305.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Survey on Distributed Digital Platform for Agricultural Supply Chain

Pratiksha Survase[1], Vaishali Kale[1], Sanjivani Durgale[1], Kshitija Gade[1], Prof. Nilesh Mali[2]
[1,2]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

Block chains are now firmly established as a digital technology that combines cryptographic, data management, networking, and incentive mechanisms to support the verification, execution, and recording of transactions between parties. While block chain technologies were originally intended to support new forms of digital currency for easier and secure payments, they now hold great promise as a new foundation for all forms of transactions. Agribusiness stands to become a key beneficiary of this technology as a platform to execute 'smart contracts' for transactions, particularly for high-value produce. First it is important to distinguish between private digital currencies and the distributed ledger and block chain technologies that underlie them. The distributed and cross-border nature of digital currencies like Bit coin means that regulation of the core protocols of these systems by central banks is unlikely to be effective. Monetary authorities are focused more on understanding 'on-ramps' and 'off-ramps' that constitute the links to the traditional payments system rather than being able to monitor and regulate the currency itself. In contrast to the digital currency feature of block chain, the distributed ledger feature has the potential for widespread use in agribusiness and trade financing, especially where workflows involve many different parties with no trusted central entity.

Keywords-Agriculture, supply chain, BCT, Cryptography, etc.

I. INTRODUCTION

An increasing demand in society for greater information about food reflects the need for more transparency and the lack of trust. At the same time, more and more food products and beverages are branded and accompanied by a variety of certification schemes, with an increasing risk of fraud (selling unqualified product with high-quality labels or claims) and adulteration. In the current situation, much of the compliance data and information is audited by trusted third parties and stored either on paper or in a centralized database and these approaches are known to suffer from many informational problems such as the high cost and inefficiency of paper-based processes and fraud, corruption and error both on paper and in IT systems. These information problems, indicating that current transparency and trust systems have not been able to solve or at times even have exacerbated the problems of low transparency and trust in agricultural food chains, pose a severe threat to food safety, food

quality, and sustainability. In particular, food integrity has become a major concern. Food integrity refers to the fairness and authenticity of food in food value chains both at the physical layer and the digital layer, where the digital layer should provide reliable and trustworthy information on the origin and provenance of food products in the physical layer. Block chain technology provides a means to ensure permanence of records and potentially to facilitate the sharing of data between disparate actors in a food value chain. This potential may lead to an exciting paradigm shift facilitating transparency and trust in food chains that ensures food integrity.

II. PROBLEM STATEMENT

1. The provenance (tracing) system of agricultural products is important for ensuring food safety.
2. However, the stakeholders (growers, farmers, sellers etc.) are numerous and physically dispersed, making it difficult to manage data and information with a centralized approach.
3. As a result, the production procedure remains non-transparent and trust is hard to build.

III. LITERATURE SURVEY

Blockchain-Based Soybean Traceability in Agricultural Supply Chain
Khaled Salah ; Nishara Nizamuddin ; Raja Jayaraman ; Mohammad Omar
Published in: IEEE Access (Volume: 7)
proposed solution eliminates the need for a trusted centralized authority, intermediaries and provides transactions records, enhancing efficiency and safety with high integrity, reliability, and security. The proposed solution focuses on the utilization of smart contracts to govern and control all interactions and transactions among all the participants involved within the supply chain ecosystem. All transactions are recorded and stored in the blockchain's immutable ledger with links to a decentralized file system (IPFS) and thus providing to all a high level of transparency and traceability into the supply chain ecosystem in a secure, trusted, reliable, and efficient manner.

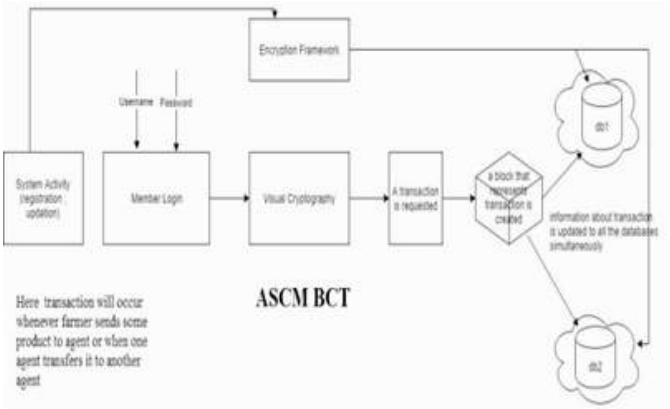
Blockchain-based traceability in Agri-Food supply chain management: A practical implementation
Miguel Pincheira Caro ; Muhammad Salek Ali ; Massimo Vecchio ; Raffaele Giaffreda
Published in: 2018 IoT Vertical and Topical Summit on Agriculture - Tuscany (IOT Tuscany)
This paper presents Agricultural Block IoT, a fully decentralized, block chain-based traceability solution for Agricultural-Food supply chain management, able to seamless integrate IoT devices producing and consuming digital data along the chain. To effectively assess Agricultural Block IoT, first, we defined a classical use-case within the given vertical domain, namely from-farm-to-fork. Then, we developed and deployed such use-case, achieving traceability using two different blockchain implementations, namely Ethereum and Hyperledger Sawtooth. Finally, we evaluated and compared the performance of both the deployments, in terms of latency, CPU, and network usage, also highlighting their main pros and cons.
Block chain Based Provenance for Agricultural Products: A Distributed Platform with Duplicated and Shared Bookkeeping
Jing Hua ; Xiujuan Wang ; Mengzhen Kang ; Haoyu Wang ; Fei-Yue Wang
Published in: 2018 IEEE Intelligent Vehicles Symposium (IV)
In this paper, we propose an agricultural provenance system based on techniques of block chain, which is featured by decentralization, collective maintenance, consensus trust and reliable data, in order to solve the

trust crisis in product supply chain. Recorded information includes the management operations (fertilizing, irrigation, etc.) with certain data structure. Applying block chain techniques to the provenance of agricultural product not only widens the application domain of block chain, but also supports building a reliable community among different stakeholders around agriculture production.

An agri-food supply chain traceability system for China based on RFID & block chain technology
Feng Tian
Published in: 2016 13th International Conference on Service Systems and Service Management (ICSSSM)
In this paper, we study the utilization and development situation of RFID (Radio-Frequency Identification) and block chain technology first, and then we analyze the advantages and disadvantages of using RFID and blockchain technology in building the agri-food supply chain traceability system; finally, we demonstrate the building process of this system. It can realize the traceability with trusted information in the entire agri-food supply chain, which would effectively guarantee the food safety, by gathering, transferring and sharing the authentic data of agri-food in production, processing, warehousing, distribution and selling links.
Blockchain application in food supply information security
Daniel Tse ; Bowen Zhang ; Yuchen Yang ; Chenli Cheng ; Haoran Mu
Published in: 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)
This article introduces the concept of Blockchain technology, putting forward the application of Blockchain technology in information security of the food supply chain and comparing it with the traditional supply chain system.

IV. PROPOSED SYSTEM

A. System Architecture:



B. Description:

Whenever any transaction will occur in the system , the record of that transaction is maintained in the form of hash value in a block. Each next block will get attached to the previous block and in this way a virtual block chain will occur. The hash value of a current block is generated using the data of a current block and the hash of the previous block. In this way if any of the block is tempered the subsequent all the block's hash must be changed . Such multiple copies are maintained at different servers , which will assure the data security and confidentiality. As everything is through application interface, it will maintain the transparency in the agricultural supply chain management.

Algorithm:

AES is used to encrypt the database.
The encryption process uses a set of specially derived keys called round keys.
These are applied, along with other operations, on an array of data that holds exactly one block of data, the data to be encrypted.
This array we call the state array.

STEPS:

1. Derive the set of round keys from the cipher key.
2. Initialize the state array with the block data (plaintext).
3. Add the initial round key to the starting state array.
4. Perform nine rounds of state manipulation.
5. Perform the tenth and final round of state manipulation
6. Copy the final state array out as the encrypted data (ciphertext).

SHA 256:

(secure hash algorithm, FIPS 182-2) is a cryptographic hash function with digest length of 256 bits. It is a keyless hash function; that is, an MDC (Manipulation Detection Code). A message is processed by blocks of 512 = 16 32 bits, each block requiring 64 rounds A cryptographic hash (sometimes called digest) is a kind of signature for a text or a data file. SHA-256 generates an almost-unique 256-bit (32-byte) signature for a text.A hash is not encryption it cannot be decrypted back to the original text (it is a one-way cryptographic function, and is a fixed size for any size of source text). This makes it suitable when it is appropriate to compare hashed versions of texts, as opposed to decrypting the text to obtain the original version.

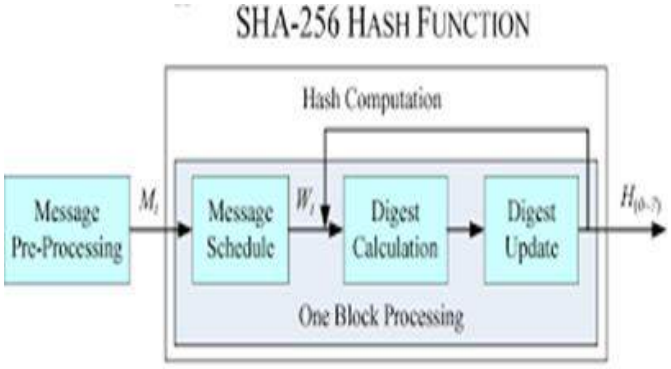


Figure 1. SHA-256 algorithm flow diagram

V. CONCLUSION

Thus we have implemented a prototype web based software application in Java for application of BCT in supply chain management . We have implemented a blockchain features such as:

1. Decentralization
2. Visual Cryptography
3. Hash Algorithm
4. Encrypted Database.

VI. REFERENCES

[1]. F. Lv and S. Chen, "Research on Establishing a Traceability System of Quality and Safety of Agricultural Products Based on Blockchain Technology," Rural Finance Research, vol. 12, pp. 22-26, 2016.

[2]. Y. Yang and Z. Jia, "Application and Challenge of Blockchain Technology in the Field of Agricultural Internet of Things," Information Technology, vol. 258, pp. 24-26, 2017.

[3]. S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," Consulted, 2008.

[4]. Y. Yuan and F. Y. Wang, "Blockchain: The State of the Art and Future Trends," Acta Automatica Sinica, 2016.

[5]. Y. Yuan, T. Zhou, A. Y. Zhou, Y. C. Duan, and F. Y. Wang, "Blockchain Technology: From Data Intelligence to Knowledge Automation," Zidonghua Xuebao/acta Automatica Sinica, vol. 43, pp. 185-1490, 2017.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Litter Detection Based on Faster R-CNN

Vishal Yadav[1], Michelle Trivedi[1], Mona[1], Jayshree Chaudhari[1], Fraz Bagwan[1]
[1]Dr. Dy Patil School of Engineering, Savitribai Phule Pune University, Pune, Maharashtra India

ABSTRACT

Cleanliness of city streets has an important impact on city environment and public health. Conventional street cleaning methods involve street sweepers going to many spots and manually confirming if the street needs to clean. However, this method takes a substantial amount of manual operations for detection and assessment of street's cleanliness which leads to a high cost for cities. Using pervasive mobile devices and AI technology, it is now possible to develop smart edge-based service system for monitoring and detecting the cleanliness of streets at scale. This paper explores an important aspect of cities - how to automatically analyse street imagery to understand the level of street litter. A vehicle equipped with smart edge station and cameras is used to collect and process street images in real time. A deep learning model is developed to detect, classify and analysis the diverse types of street litters such as tree branches, leaves, bottles and so on. In addition, two case studies are reported to show its strong potential and effectiveness in smart city systems.

Keywords: Smart City, Street Cleanliness.

I. INTRODUCTION

Urban street surface receives waste deposits from both natural and human sources, such as leaves, soil, sediment, scattered trash, illegal dumpling and so on. When the street cleaning service is ineffective, it could cause a negatively impact on city tourism, reputation, and economy. Furthermore, dirty street have also been recognized as potentially important contributor to air and water pollution. Researches have proved that if there is litter on the streets, people do not hesitate in throwing more litter.
However, if the streets are clean, people tend to think twice before throwing anything and end up not throwing the litter on streets. Therefore, cleanliness of city streets is particularly important as it has a significant effect on city's image and reputation, and on the quality of life for those who live and work in the city.
Currently, many cities have adopted various methods and made great effort to improve the cleanliness of their streets. For instance, New York city used an inspection program called Scorecard to

measure the cleanliness of city streets and sidewalks [23]. [20] proposed a cleanliness index for the city of Granada (South of Spain) to measure the level of cleanliness of the streets. In 2015, Imteaj et al. constructed an android based application for the city Dhaka, capital of Bangladesh. The user himself can

contribute to clean his city, notify volunteer to come forward or can inform city corporation. In 2015, Los Angeles had developed a state-of-the-art street-by-street [12] cleanliness assessment system. Los Angeles is the first city to map the cleanliness of every one of its blocks. With this new tool, Sanitation is better equipped to target areas of high need, and to ensure a fair allocation of services.

Despite these methods provided new ideas for cities to clean their streets, most of the current methods for detection litter are not fully automated and still rely on human intervention. The clean-up crews need to capture and identify each picture manually to determine if the street is dirty. Therefore, a promising and optimal solution should automatically and reliably detect litter in each captured image without human intervention.

The major motivation of this paper is providing cities with an automated way to monitor the street cleanliness. In this paper, we have studied a smart clean street service application using the state-of-art advanced technologies in IoT, mobile edge computing, big data analysis, as well as machine learning techniques. The mobile edge processing component located in the vehicle which carries the cameras. It shall be able to process the images using a limited computing power and send those images to the service server which may benefit from deeper scan using the powerful GPU processing on the server. After the images pre- processing at the edge layer, Faster R-CNN (Faster Region- Convolutional Neural Network) is used to

identify the street litter at the service server. Finally, the analysis results are sent to user layer for evaluation.

The main contribution of this paper can be summarized as follows. First, we constructed a unique litter data set of 12346 images for training deep-learning based detection models. The second contribution is a litter detector which can detect 11 classes of litter with good accuracy. The third contribution is an edge computing framework of deep learning for smart city street cleanliness.

The rest of the paper is organized as follows. In Section 2, we introduce related work in smart street cleaning, deep learning, and edge computing. In Section 3, we present the architecture, components, and algorithms for the proposed system based on deep learning and edge computing. In Section 4, we describe the implementation details of our system. Section 5 presents the evaluation results. Finally, in Section 6, we make concluding remarks.

II. Motivation

Keep India Beautiful aims to create a cleaner India and hosts community clean up events across the country. However, their events are done without any strong statistical data. With the data which could be taken from our algorithm, they will be able to identify the locations in the India that needs to be cleaned up the most.

Clean India will lead to decrease in the number of deceases caused by unhygienic atmosphere.

The major contributions of this paper are presented as follows.

- We studied a Faster R-CNN open source framework with region proposal network and Resnet network algorithm, using ResNet network to replace the previous VGG network as the basic convolution layers.
- To optimize the performance of the model, we collect urban scene images containing garbage and urban scene images without garbage. By using fine-tuning strategy, we apply the pre-training model parameters which has been trained in coco dataset to our network.
- We propose a dataset fusion strategy, which integrates the garbage dataset with several other

datasets of typical categories in urban scenes. In summary, the method has near- real time and generalization capabilities. Through experiments, we observe that the false detection rate of the garbage area y

III. Related Work

This section reviews the existing research work and related projects.

Smart street cleaning:

The cleanliness of city street is directly related to the city's public image. To maintain the streets clean, different methodologies have been developed in the past years. These methodologies can be classified into two directions: evaluating the street cleanliness, monitoring the waste. In order to evaluate the street cleanliness, Seville etal.proposed a clean index for measuring the level of cleanliness of the city streets, such that the quality and governance of public services can be evaluated. However, the process of measurement requires a lot of human intervention like collecting data and rating data. Lopez et al. develop an App to evaluate the street cleanliness and waste collection service. Specific methods for calculating and evaluating indicators have been designed to give a true reflection of the level of city street cleanliness.

Although this App can collect information from the user end and store information in the application database, it still needs users to fill the information manually in the App. Li et al. put forward a multi-level assessment system and showed how the cleanliness status of streets is collected by using mobile stations. The results are transmitted through city network, analysed in the cloud and presented to city administrators online or on mobile. Regarding monitoring the waste, Rovetta . used sensors to monitor waste bins based on distributed sensor technology and geographical information systems. Begur et al.[3] focused on illegal dumping problems in the City of San Jose. They proposed an innovative smart mobile-based service system, which supports real-time illegal dumping detection, altering, monitoring, and management. Alfarrarjeh et al. [1] presented an automating geo-spatial classification approach to determine the level of street cleanliness. The experiments compared various combinations of classifier and image features, which show that SVM classifier based on CNN image features obtained good values on both precision and recall.

Balchandani et al. [2] proposed a deep learning framework for smart street cleaning, which aims at providing any city with an automated way to monitor the cleanliness of its streets. It is a good idea to use deep leaning technology to automatically detect and classify litter, but this paper only provided with a simple example about separating the street and the curb, and the performance of detection and classification was not discussed. The proposed approach in this paper is also based on recent advances in deep learning.

Related work in deep learning is introduced in the next paragraph.

IV. System Design

This section presents the overall infrastructure of the system. All modules are explained in detail.

System Infrastructure:

Due to the recent advances of IoT technologies, mobile edge computing, big data analysis, as well as machine learning, we designed the system architecture of the proposed smart clean street service system, depicted in Fig. 1. It shows that the

system architecture includes three layers: edge layer, cloud layer and user layer. First, images of streets are captured on a mobile device and pre-processed at the edge layer. Also, the end mobile device could record the geological location and provide location awareness for end users. Then, these pre-processed images with location information will be sent from

the edge layer to cloud layer. The cloud-based server performs additional processing on the incoming images and then pushed them through a deep learning algorithm for object detection and classification. Finally, the detail results including litter location, types of litter, and litter detection photos are fed into the end user database for visualization and reporting.

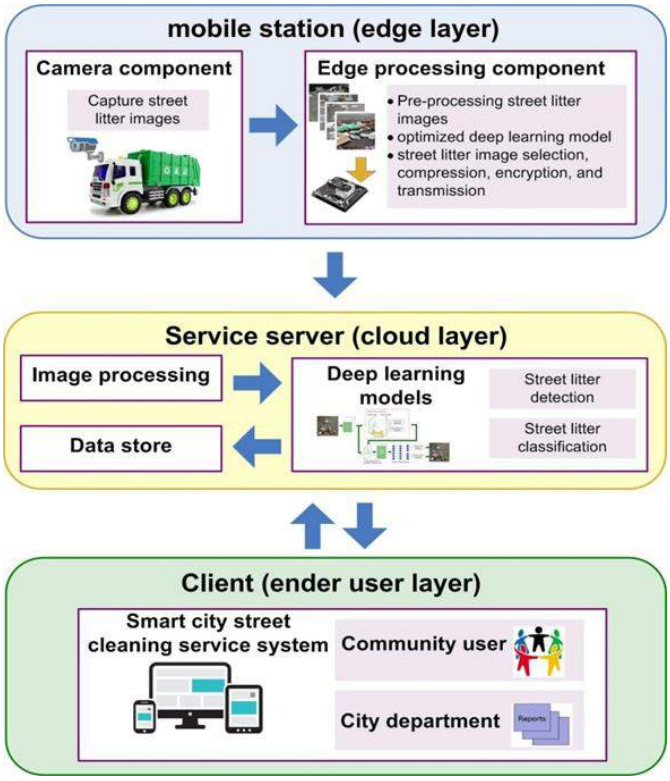


Figure 1: Clean Streets Framework -System Architecture [24]

Edge layer: The main goal of this layer is to capture the street images and collect location coordinates from streets using a smart vehicle with a mobile station. This mobile station has two major components: Camera module and Edge processing module. Three cameras in different direction takes high-resolution pictures of the street. Each camera has a predefined angle and resolution which can be customized and covers defined range 20ft. All these captured pictures are completely in raw or uncompressed, and passed into the edge processing module. The edge processing module is made of several parts. First, an optimized pre-trained deep learning model is used to determine any regions of interest in images which are worth deeper processing in the server, and to check whether the image contains a clear view of the street and is not blur. Then, the selected images are compressed,

encrypted, and finally transmitted to the server through the wireless network.

Server layer: This is the layer where the pre-processed images are further analysed on the cloud server, which is configured to use Tensor flow (an open source software interface) for CNN model training and testing. First, the server-side image processor decrypt the image. Then, the decrypted image is fed into the pre-trained deep learning models like Faster-RCNN which detect and classify the object in the image. Lastly, results generated by the object detection and classification are fed into the end user database for visualization and reporting. User layer: This is the layer where reports are generated based on the Cloud processing. These results including litter location, types of litter, and litter detection photos are visualized for city and community.

Network architectures for litter recognition:

1) Faster R-CNN

In this section, we introduce the key aspects of the Faster R-CNN. [18] Faster R- CNN uses CNN layers named Region Proposal Network (RPN) instead of Selective Search. Faster R-CNN generates region

proposals from feature maps generated by RPN. RPN scans the sliding window on the feature map and extracts object candidates.[10] At this time, in order to detect long slender objects, each grid have some bounding boxes called anchor boxes. Finally, like the Fast R- CNN, the region proposals are projected onto the feature map, and the object is detected by classifying region proposals. Since Faster RCNN generates image features and region proposals using single CNN, there is an advantage that end-to-end training can be performed in addition to faster detection

V. CONCLUSION

This paper presents a deep learning based smart street cleaning service system that allows for automatic litter detection and classification and real- time monitoring of the streets conditions. In the proposal system, we studied deep learning model for little street object detection and classification. We have successfully learned and studied python language and have also acquired knowing about libraries like tensorflow and have also learned about object detection and image classification algorithms.

edge-based service system for street cleanliness assessment.

VI. REFERENCES

[1]. Alfarrarjeh, A., Kim, S. H., Agrawal, S., Ashok, M., Kim, S. Y., &Shahabi, C. (2018). Image classification to determine the level of street cleanliness: A case study. In 2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM) (pp. 1–5).

[2]. Balchandani, C., Hatwar, R. K., Makkar, P., Shah, Y., Yelure, P., &Eirinaki, M. (2017). A deep learning framework for smart street cleaning. In 2017 IEEE Third International Conference on Big Data Computing Service and Applications (BigDataService) (pp. 112–117).

[3]. Begur, H., Dhawade, M., Gaur, N., Dureja, P., Gao, J., Mahmoud, M., Huang, J., Chen, S., & Ding, X. (2017). An edge-based smart mobile service system for illegal dumping detection and monitoring in san jose. In 2017 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computed, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI) (pp. 1–6). doi:10.1109/UIC-ATC.2017.8397575.

[4]. Gertler, A., Kuhns, H., Abu-Allaban, M., Damm, C., Gillies, J., Etyemezian, V., Clayton, R., & Proffitt, D. (2006). A case study of the impact of winter road sand/salt and street sweeping on road dust re-entrainment. Atmospheric Environment, 40, 5976 – 5985.

[5]. Girshick, R. (2015). Fast r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 1440–1448).

[6]. Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580–587).

[7]. He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 37, 1904–1916.

[8]. Huang, L.-P., Hong, M.-H., Luo, C.-H., Mahajan, S., & Chen, L.-J. (2018). A vector mosquitoes classification system based on edge computing and deep learning. In 2018 Conference on Technologies and Applications of Artificial Intelligence (TAAI) (pp. 24–27). IEEE.

[9]. Imteaj, A., Chowdhury, M., &Mahamud, M. A. (2015). Dissipation of waste using dynamic perception and alarming system: A smart city application. In 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT) (pp. 1–5). doi:10.1109/ICEEICT.2015.7307410.

[10]. Iriarte, A., Gabarrell, X., &Rieradevall, J. (2009). Lca of selective waste collection systems in dense urban areas. Waste Management, 29, 903 – 914.

[11]. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097–1105).

[12]. Ladin-Sienne, S. (2017-04-16). Turning dirty streets clean through comprehensive open data mapping. URL: https://datasmart.ash. harvard.edu/news/article/turning-dirty-streets-clean-through-comprehensive-open-data-mapping-1001.

[13]. Li, H., Ota, K., & Dong, M. (2018). Learning iot in edge: deep learning for the internet of things with edge computing. IEEE Network, 32, 96–101.

[14]. Li, W., Bhushan, B., Gao, J., & Zhang, P. (2018). Smartclean: Smart city street cleanliness system using multi-level assessment model. International Journal of Software Engineering and Knowledge Engineering, 28, 1755 – 1774.

[15]. Liu, C., Cao, Y., Luo, Y., Chen, G., Vokkarane, V., Yunsheng, M., Chen, S., & Hou, P. (2018). A new deep learning-based food recognition system for dietary assessment on an edge computing service infrastructure. IEEE transactions on services computing, 11, 249–261.

[16]. Lpez, I., Gutirrez, V., Collantes, F., Gil, D., Revilla, R., & Gil, J. L. (2017). Developing an indicators plan and software for evaluating street cleanliness and waste collection services. Journal of Urban Management, 6, 66 – 79.

[17]. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91–99).

[18]. Ren,S.,He,K.,Girshick,R.,& Sun,J. (2015).Faster r-cnn:Towards real-time object detection with region proposalnetworks.InC.Cortes,N.D.Lawrence,D. D. Lee, M.Sugiyama,&R.Garnett(Eds.),AdvancesinNeuralInformationProcessing Systems28(pp.91–99).CurranAssociates,Inc.URL: http://papers.nips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks.pdf.

[19]. Rovetta, A., Xiumin, F., Vicentini, F., Minghua, Z., Giusti, A., &Qichang, H. (2009). Early detection and evaluation of waste through sensorized containers for a collection monitoring application. Waste Management, 29, 2939 – 2949.

[20]. Sevilla, A., Rodrłguez, M. L., ¢ngelaGarcła-Maraver, &Zamorano, M. (2013). An index to quantify street cleanliness: The case of granada (spain). Waste Management, 33, 1037 – 1046.

[21]. Silva, B. N., Khan, M., & Han, K. (2018). Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities. Sustainable Cities and Society, 38, 697 – 713. URL: http://www.sciencedirect.com/science/article/pii/S2210670717311125.doi:https://doi.org/10.1016/j.scs.2018.01.053.

[22]. Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., & Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE transactions on medical imaging, 35, 1299–1312.

[23]. Van Ryzin, G., Immerwahr, S., & Altman, S. (2008). Measuring street cleanliness: A comparison of new york city's scorecard and results from a citizen survey. Public Administration Review, 68, 295 – 303.

[24]. Smart street litter detection and classification based on Faster R-CNN and edge computing - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Clean-Streets-Framework- System-Architecture_fig1_338988152 accessed 23 Nov, 2020]

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Design and Implementing Brain tumor Detection Using Machine Learning Approach

Swati Jagtap[1], Sadichha Khedkar[1], Meghana Rikibe[1], Sampada Pathare[1], Prof. Amruta Chitari[2]
[1,2]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

In this paper, we propose a brain tumor segmentation and classification method for multi-modality magnetic resonance image scans. The data from multi-modal brain tumor segmentation challenge are utilized which are co-registered and skull stripped, and the histogram matching is performed with a reference volume of high contrast. We are detecting tumor by using preprocessing , segmentation, feature extraction ,optimization and lastly classification after that preprocessed images use to classify the tissue .We performing a leave-oneout cross-validation and achieved 88 Dice overlap for the complete tumor region, 75 for the core tumor region and 95 for enhancing tumor region, which is higher than the Dice overlap reported.

I. INTRODUCTION

The detection and diagnosis of brain tumor from MRI is crucial to decrease the rate of casualties. Brain tumor is difficult to cure, because the brain has a very complex structure and the tissues are interconnected with each other in a complicated manner. Despite many existing approaches, robust and efficient segmentation of brain tumor is still an important and challenging task. Tumor segmentation and classification is a challenging task, because tumors vary in shape, appearance and location. It is hard to fully segment and classify brain tumor from mono-modality scans, because of its complicated structure. MRI provides the ability to capture multiple images known as multimodality images, which can provide the detailed structure of brain to efficiently classify the brain tumor. shows different MRI modalities of brain.

To design a detection and diagnosis of brain tumor from MRI is crucial to decrease the rate of casualties. Brain tumor is difficult to cure, because the brain has a very complex structure and the tissues are interconnected with each other in a complicated manner. Despite many existing approaches, robust and efficient segmentation of brain tumor is still an important and challenging task. Tumor segmentation and classification is a challenging task, because tumors vary in shape, appearance and location. It is hard to fully segment and classify brain tumor from mono-modality scans, because of its complicated structure. So we overcome that problem classify the brain tissues tumor area.

Robust and efficient segmentation of brain tumor is still an important and challenging task. Tumor segmentation and classification is a challenging task,

because tumors vary in shape, appearance and location. It is hard to fully segment and classify brain tumor from mono-modality scans, because of its complicated structure. So we overcome that problem classify the brain tissues tumor area.

The brain images taken as input and that images performs the preprocessing operation after the preprocessing segmentation using the k-means algorithm and on that segmented area we perform the operation feature extraction using the classification SVM and CNN algorithm.

This technique can be developed to classify the tumors based on feature extraction. the detection of brain tumor is fast and accurate

when compared to the manual detection carried out by clinical experts.

II. Methodology

- Support Vector Machine The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space(N—the number of features) that distinctly classifies the data points.

- Possible hyperplanes To separate the two classes of data points, there are many possible hyperplanes that could be chosen. Our objective is to find a plane that has the maximum margin,
  i.e the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence.

- Hyperplanes in 2D and 3D feature space Hyperplanes are decision boundaries that help classify the data points. Data points falling on either side of the hyperplane can be attributed to different classes. Also, the dimension of the hyperplane depends upon the number features. If the number of input features is 2, then the hyperplane is just a line. If the number of input features is 3, then the hyperplane becomes a two-dimensional plane. It becomes difficult to imagine when the number of features exceeds 3.

- Support Vectors Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane. Using these support vectors, we maximize the margin of the classifier. Deleting the support vectors will change the position of the hyperplane. These are the points that help us build our SVM.

- Large Margin Intuition In logistic regression, we take the output of the linear function and squash the value within the range of [0,1] using the sigmoid function. If the squashed value is greater than a threshold value(0.5) we assign it a label 1, else we assign it a label 0. In SVM, we take the output of the linear function and if that output is greater than 1, we identify it with one class and if the output is -1, we identify is with another class. Since the threshold values are changed to 1 and -1 in SVM, we obtain this reinforcement range of values([-1,1]) which acts as margin.

- CNN (Convolutional Neural Network)

- Mathematical Working and Flow of CNN:

[1] Input Layer — has input size [7, 1] because we have 7 features
[2] conv1d— First convolutional layer
[3] averagePooling1d — First average pooling layer
[4] conv1d — Second convolutional layer
[5] averagePooling1d — Second pooling layer
[6] flatten — Reduce the dimension, reshape input to [number of samples, number of features]
[7] dense — Fully connected layer using linear activation function with 1 unit which returns 1 output value

- Convolutional neural network (CNN) is part of the family of neural network (NN) which is a variation of a multilayer perceptron (MLP). CNN consists of an input layer, several hidden layers and an output layer like any other NNs. Input layer is a representation of identity function, f(x) = x. Output layer which makes decisions, passes previously calculated weights through a linear function. Hidden layers are either convolutional, pooling, dropout or fully connected. In addition, all layers have activation functions at the end which gives additional functionality
  e.g. normalization. sigmoid , tanh and RELU are examples of these activation functions. Weights of convolution layers can be seen as 2D- filters and they apply convolution operation with these filters. Convolution operation is a process which sums the point-wise multiplications of given two functions while sliding the operation window. Pooling layer generalizes the elements in window frame while sliding this window. For example, max pooling outputs the maximum elements for a given window while sliding it. Dropout selects several neurons, that feed the input of next layers and reduces overfitting. Finally, fully connected layers can be thought as a fully connected version of classical MLP. With the explanation of fully connected layers, CNNs can be seen as a combination of MLP and filters which can operate as either convolution or pooling. In order to optimize weights of CNN, we have used an adaptive learning rate method (ADADELTA) optimizer. Apart from the fact that CNNs give noteworthy performance , they require much more data compared to other types of models. With the purpose of solving this issue, we merge all ETFs and create a satisfactory dataset for a financerelated problem. Additionally, merging reveals that we do not have to stick to one set of stock only. The model trained with this dataset, assumed that it performs well, will be universal for all kinds of stocks rather than a particular stock.

$$\tanh(x) = 1\ e2x/\ 1 + e2x \quad (1)$$
$$(x) = 1\ 1 + ex \quad (2)$$
$$R(x) = \max(0, x) \quad (3)$$

$$yi = 01 + 1xi1 + \ldots + pxip = xT\ i\ ,\ i = 1, \ldots, n \quad (4)$$

where T denotes the transpose and xT i is the inner product between two vector.

A. Project outline

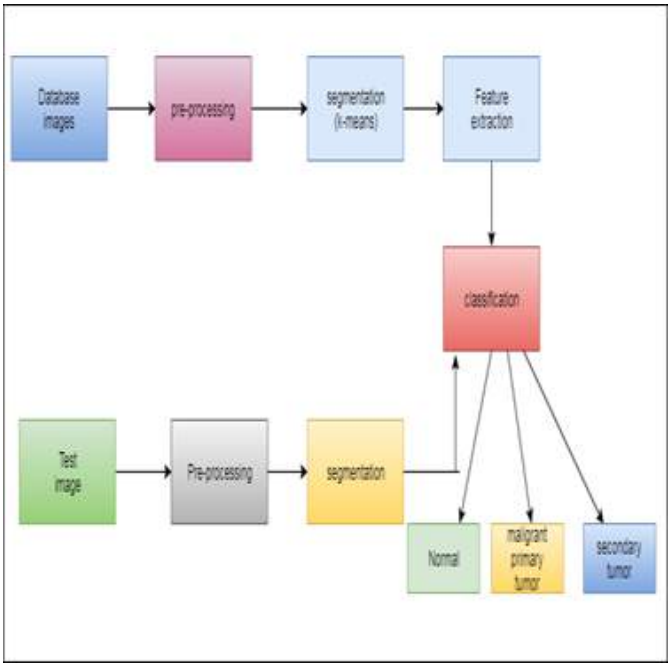


Figure 1: Overall Architecture

B. Brain image Preprocessing

Due to existing noise disturbance the MR images get affected. For noise reduction the research work proposes local smoothing methods and nonlocal means. In the image there may also exist few significant structures and details that can act as noise; such kind of details are also eliminated. The technique of Image pre-processing involves: data cleaning, data transformation, data integration, data resizing, data reduction etc. The image pre-processing eliminates unnecessary data and smooth up noisy data, detect and eliminate the outlier and rectify the data inconsistencies. Lastly, normalization and aggregation is performed. The technique of Image-processing proves to be highly significant in determining particular heart image, removing noise and for improvising the quality of the image

C. Average filtering

The normal channel being the convolution work that is utilized to set the clamor in the images. The Pre-processing step abandons the disturbances in the image but still after applying preprocessing the image doesn't hold suitable for future process. As a result the Average channel resolves this issue by

providing acceptable and smooth picture. The Average channel resembles a non-linear channel unlike straight channels. The Average channel replace the pixel esteems with an Average esteem that being nearly accessible (like, 3x3 or 5x5 or pixels near the focal pixel esteem). Moreover, Average channel tends to be edge safeguarding. It helps in abandoning salt and pepper disorder.

Algorithm:

Step 1 : The picture is provided as input.
Step 2 : Choose a 3X3 window near the current pixelwithin the picture.
Step 3 : Perform pixel sorting in expanding request andsave it to vector.
Step 4 : Determine the normal of the vector.
Step 5 : The current pixel is replaced with the normalesteem.
Step 6 : Repetition of means 2 to 5 till every single pixelswithin the picture gets prepared.
Step 7 : Output.

D. Pixel based segmentation

Image Segmentation is a common technique of digital image processing. Lately, Brain tumor image sectioning in MRI has spurred up as a popular research in the domain of medical imaging system. The process of Segmentation

E. Convolution Neural Networks

Convolutional Neural Network – CNN is employed for segmenting the images. It directly extracts features from pixel images with least pre-processing involved. The network utilized is LinkNet which being a light deep neural network architecture that's developed to carry out semantic segmentation. The LinkNet Network contains encoder and decoder blocks which basically manage to split the image and re-build again before it's forwarded via few final convolutional layers. CNN is a significant approach of deep learning which is being employed in image recognition applications. It involves two basic methods of

convolution and pooling. Convolution and pooling layers are arranged till high level of classification accuracy is achieved. Moreover, few feature maps are

identified in every convolutional layer and weights linked to convolutional nodes (in the same map) are being shared. Such arrangements offer comprehension of various network characteristics at the same time retaining the no: of traceable parameters. CNN possess less specific tasks in contrast to the conventional methods and helps in thoroughly extracting features. Figure 2, depicts the CNN process scheme as.

Algorithm for CNN based Classification

Step 1 : Convolution filter is applied in the first layer.

Step 2 : The filter sensitivity is minimized by smoothingthe convolution filter that is by sub- sampling.

Step 3 : The activation layer controls the signal transferfrom one layer to other layer.

Step 4 : Training period is being fastened by employingRELU (rectified linear unit).

Step 5 : The neurons in proceeding layer is associated witheach neuron in the next layer.

Step 6 : At the time of training, Loss layer is appended inthe end to provide a feedback to NN (neural network).

F. Evolution metrics

For performance evaluation and measuring system stability, few parameters are computed and examined. These are mentioned as:

The proposed CNNs performance is assessed with RMSE(Root Mean Square Error), recall, sensitivity, precision, F-score specificity, PME (probability of the misclassification error) and accuracy of the training and testing set and throughout performance was examined by making use of the Eqs. (1-8) correspondingly, where Y_i

denotes actual and R_i denotes result of the i^{th} diagnosis of brain tumor feature acquired, TN (True Negative) denotes prediction for the patients with no brain tumor and were detected with no brain tumor, FN (False Negative) denotes the prediction for the patients with no brain tumor but were detected with a brain tumor, TP(True Positive) denotes the prediction for the patients with brain tumor and were detected with a brain tumor, and FP(False Positive) represents the prediction for the patients having brain tumor but were detected with no brain tumor.

- True Positive (TP): If the instance is positive and it is classified as positive
- False Negative (FN): If the instance is positive but it is classified as
- True Negative (TN): If the instance is negative and it is classified as negative
- False Positive (FP): If the instance is negative but it is classified as positive.

III. LITERATURE RIVIEW

Sr no	Title	Authors	Year	Description
1	Semi-Automatic Segmentation Software for Quantitative Clinical Brain Glioblastoma Evaluation	Ying Zhu, MS, Geoffrey S. Young, MD, Zhong Xue, PhD, Raymond Y. Huang, MD, PhD, Hui You	2011	Our software adopts the current state-of-the-art tumor segmentation algorithms and combines them into one clinically usable pipeline.
2	A novel content-based active contour model for brain tumor segmentation	Sachdeva J, Kumar V, Gupta I, Khandelwal N, Ahuja CK	2011	Brain tumor segmentation is a crucial step in surgical and treatment planning. Intensity-based active contour models such as gradient vector flow (GVF), magneto static active contour (MAC) and fluid vector flow (FVF) have been proposed to segment homogeneous objects/ tumors in medical images.
3	New developments in liposomal	B.S.	20	Five decades of research in the

	drug delivery	Pattni, V.V. Chupin and V.P. Torchilin	14	Field of liposome research have shown their prospective benefits in the medical and cosmetic as well as the food industry.
4	Liposomal drug delivery systems: From concept to clinical applications	T.M. Allen and P.R. Torchilin	20 14	The first closed bilayer phospholipid systems, called liposomes, were described in 1965 and soon were proposed as drug delivery systems.

IV. CONCLUSION

In this paper we are presenting an algorithm to hierarchically classify the tumor into three regions: whole tumor, core tumor and enhancing tumor. Intensity, intensity difference, neighborhood information and wavelet features will be extracted and utilized on multi-modality MRI scans with various classifiers. The use of SVM and CNN classifier will increase the classification accuracy as evident by quantitative results of our proposed method which are comparable or higher thanthe state of the art.

V. REFERENCES

[1]. S. Yazdani, R. Yusof, A. Karimian, Pashna, and A. Hematian, "Image segmentation methods and applications in MRI brain images," IETE Technical Review, vol. 32, no. 6, pp. 413–427, 2015.

[2]. M. P. Arakeri and G. R. M. Reddy, "Computer-aided diagnosis system for tissue characterization of brain tumor on magnetic resonance images," Signal, Image and Video Processing, vol. 9, no. 2, pp. 409–425, 2015.

[3]. E. Dandil, M. Cakiroglu, Z. Eksi, "Computer-aided diagnosis of malign and benign brain tumors on MR images," ICT Innovations 2014, pp. 157–166, 2015.

[4]. E.-S. A. El-Dahshan , H. M. Mohsen, K. Revett, and A.-B. M. Salem, "Computer- aided diagnosis of human brain tumor through MRI: A survey and a new algorithm," Expert Systems with Applications, vol. 41, no. 11, pp. 5526– 5545, 2014.

[5]. Kailash Sinha, G.R.Sinha, Efficient Segmentation Methods for Tumor Detection in MRI Images", IEEE Student's Conference on Electrical, Electronics and Computer Science, pp 1-6,2014

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Data Security in Cloud Computing

Rohit Hodage[1], Ritesh Hajare[1], Om Wangwad[1], Ashwini Pandagale[2], Faraz Bagwan[2]

[1]Computer Engg. Savitribai Phule, Pune University, Pune, Maharashtra, India

[2] Assistant Professor, Computer Engg. Savitribai Phule, Pune University, Pune, Maharashtra, India

ABSTRACT

Data security has been consistent in being a major issue in information technology. In the cloud computing world, it becomes specifically critical as the data is situated in different places all over the world.

As per user's concerns about the cloud technology the important factors are privacy protection and data security. In both academics and industries, the topics in cloud computing have been checked by multiple techniques. For the future growth of cloud computing technology in industry, government and business the data security and privacy protection will become more crucial.

Data security and privacy protection challenges are similar to both hardware and software in the cloud architecture. This study is to analyze different security techniques and challenges from both software and hardware aspects to secure data in the cloud and focuses on improving the data security and privacy protection for the trust worthy cloud environment. In this document, we are preparing a relevant research analysis on the existing research work with reference to the data security and privacy protection techniques of cloud computing.

Keywords: Data security, Privacy Protection, Cloud Computing,

I. INTRODUCTION

Cloud computing has been emerged as the next generation paradigm in computation. In the cloud computing world, both applications and resources are delivered on demand over the Internet as services. Cloud is an environment of the hardware and software resources in the data centers that provide diverse services over the network or the Internet to fulfill user's requirements .The explanation of "cloud computing" as per the National Institute of Standards and Technology (NIST) states that cloud computing allows unique, convenient. Network access to a shared pool of configurable computing resources like servers, networks, applications, storage, and services and can be provisioned on priority and released with less management effort or service provider interaction on demand.

As per the description, cloud computing provides a convenient on-demand network access to a shared pool of configurable computing resources. Resources are identical to computing applications, network resources, platforms, software services, virtual servers, and computing infrastructure[2]

Cloud computing can be looked upon as a new computing archetype that can provide services on demand at a cheap cost. The three renowned and

mostly used service models in the cloud paradigm are software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS). In SaaS, software with the similar data is deployed by a cloud service provider, and users can access it through the web browsers [1]. In PaaS, a service provider delivers services to the users with a set of software programs that can resolve the given tasks. In IaaS, the cloud service provider facilitates services to the users with virtual machines and storage to enhance their business capabilities. Cloud computing is quite similar to but not the same as grid computing.

Grid computing integrates diverse resources together and manages the resources with the unified operating systems to provide better performance computing services, while cloud computing is the combination of computing and storage resources handled by different operating systems to provide services such as large-scaled data storage and top performance computing to users. The overall picture of grid computing has been replaced by cloud computing. Distribution of data is in a new format of cloud computing comparing with the grid computing[5].

Cloud computing will enable services make the services accessible to be consumed easily on demand. Cloud computing has the features like on-demand self-service, ubiquitous network access, location independent resource pooling, rapid resource elasticity, usage-based pricing, and transference of risk. These achievements of cloud computing have attracted substantial interests from both the industrial world and the academic research world.

Cloud computing technology is currently changing the business strategy in the world. Cloud computing is very promising for the IT applications; however, there are still some concerns to be resolved for personal users and enterprises for data storage and deploy applications in the cloud computing environment. One of the most significant obstacle to adoption is data security, which is accompanied by issues including compliance, privacy, trust, and legal matters[4].

The characteristics of institutions and institutional evolution is very similar to privacy and security in cloud computing .Data security has been consistent in being a major issue in IT. Data security turns out to be very critical in the cloud computing environment, because data gets scattered in different machines and storage devices including servers, PCs, and different mobile devices like wireless sensor networks and smart phones. Data security in the cloud computing is much more complex than data security in the traditional information systems [7].

To make the cloud computing adaptive by users and enterprise, the security challenges of users should be resolved first to make cloud world trustworthy. The trustworthy environment is the basic prerequisite to win confidence of users to get used to such a technology. Discussed the assessment of cloud computing risks. Before the data security concerns are disclosed, the functions of cloud computing are reviewed first. Cloud computing is also known as on-demand service. In the cloud computing world, there is a cloud service provider that facilitates and manages the services.

The cloud provider facilitates all the services over the Internet, while end users use services for meeting their business needs and then pay for the services accordingly. Cloud computing world enables two basic types of functions which includes computing and data storage. In the cloud computing environment, users of cloud services don’t need anything and they can get access to their data and complete their computing tasks just through the Internet devices. During the access to the data and computing, the clients don't even know where the

data gets stored and which machines performs the computing tasks[2].

Coming to data storage, data safety and security are the primary factors for gaining user's trust and making the cloud technology successfully used. Many data protections and data security techniques have been launched in the research world of cloud computing. However, data security related techniques need to be further improved. Services of cloud computing has been delivered across the entire computing spectrum. Nowadays, organizations and companies are moving and expanding their business by adopting the cloud computing reduce the cost. This can be a contribution to free more man-powers to focus on developing strategic differentiation and business division of labor is transparent [10].

The concept of cloud has many implementations based on the services from service providers. For example, Google Apps Engine, Microsoft Azure, and Amazon Stack are famous implementations of cloud computing provided by cloud service providers like Google, Microsoft, and Amazon companies. Apart from the rest, the ACME enterprise implemented VMware based v-Cloud for allowing multiple organizations to share computing resources[2]

As per the difference of access scope, cloud can be distributed into three segments: public cloud, private cloud, and hybrid cloud. Public cloud has the similar property of service provider and is accessible in public, private cloud refers to being the property of a company, and hybrid cloud is the blends of public and private cloud. Most of the existing cloud services are delivered by large cloud service companies such as Google, Amazon, and IBM.

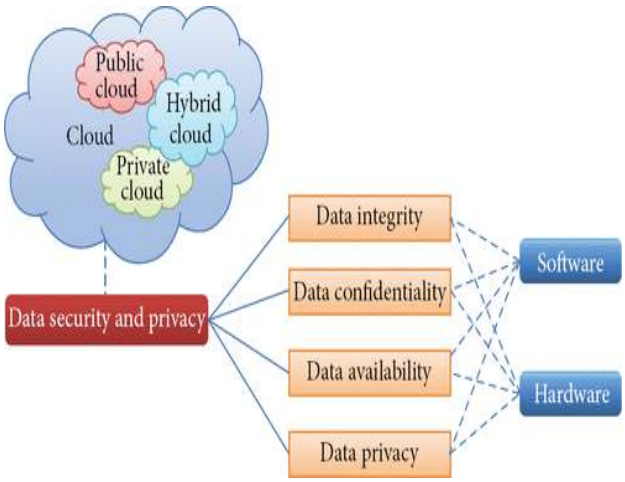
A private cloud is a cloud in which only the authorized users have access to the services from the provider. In the public cloud anybody can access the cloud services whereas the hybrid cloud contains the concept of both public and private clouds. Cloud computing can reduce the cost and save organizations time, but trusting the system is more vital because the actual asset of any organization is the data which they share in the cloud to use the required services by putting it either directly or in the relational database or eventually in a relational database using an application[4].

Cloud computing brings multiple attributes that requires serious attention when it comes to trusting the system. The trust of the entire system relies on the data protection and prevention techniques used in it. By the researchers numerous different techniques and tools have been introduced and tested for data protection and prevention to gain and remove the hurdle of trust but there are still gaps which need attention and are required to be lined up by making these techniques much better and effective. The meaning of security is plentiful. Security is the combination of confidentiality, the prevention of the unauthorized disclosure of information, integrity, the prevention of the unauthorized amendment or deletion of information, and availability, the prevention of unauthorized withholding of information [9].

The resource monitoring, resource management and resource security are the major issues in cloud computing. Currently, to deploy there are no regulations and standard rules. Applications in the cloud and there is a lack of standardization control in the cloud. In cloud numerous novel techniques had been implemented and designed; however, due to the dynamics of the cloud environment these techniques fall short of ensuring total security[12]

II. RESULTS AND DISCUSSION:

3.1 Figures and Tables:



III.CONCLUSION

Cloud computing is a promising and emerging technology for the next generation of IT applications. The barrier and hurdles toward the rapid growth of cloud computing are data security and privacy issues. Reducing data storage and processing cost is a mandatory requirement of any organization, while analysis of data and information is always the most important tasks in all the organizations for decision making. So no organizations will transfer their data or information to the cloud until the trust is built between the cloud service providers and consumers.

A number of techniques have been proposed by researchers for data protection and to attain highest level of data security in the cloud. However, there are still many gaps to be filled by making these techniques more effective. More work is required in the area of cloud computing to make it acceptable by the cloud service consumers. This paper surveyed different techniques about data security and privacy, focusing on the data storage and use in the cloud, for data protection in the cloud computing environments to build trust between cloud service providers and consumers .

IV. REFERENCES

[1]. "Forbes: Cloud computing forecast," https://www.forbes.com/sites/louiscolumbus/2017/04/29/roundup-of-cloud-computing-forecasts2017/#5c42322c31e8/, 2020.

[2]. "Microsoft onedrive,"https://products.office.com/en-us/onedrive/online-cloud-storage, 2020.

[3]. C. Wang, N. Cao, J. Li, K. Ren, and W. Lou "Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data", IEEE Transactions on Parallel and Distributed Systems, vol.23, No.8, Aug.2012.

[4]. S. Karen, "Iot big data security and privacy versus innovation," IEEE Internet of Things Journal, vol. 6, no. 2, pp. 1628 – 1635, 2019.

[5]. Z. Lei, F. Anmin, Y. Shui, S. Mang, , and K. Boyu, "Data integrity verification of the outsourced big data in the cloud environment: A survey," Journal of Network and Computer Applications, vol. 112, pp. 1–15, 2019.

[6]. T. Ye, X. Peng, and J. Hai, "Secure data sharing and search for cloud-edge-collaborative storage," IEEE Access, vol. 7, pp. 15 963 – 15 972, 2019

[7]. V. Chang and M. Ramachandran, "Towards achieving data security with the cloud computing adoption framework," IEEE Transactions on Services Computing, vol. 9, no. 1, pp. 138–151, Jan 2016.

[8]. J. Wei, W. Liu and X. Hu, "Secure and efficient attribute-based access control for multi authority cloud storage", IEEE Syst. J., vol. 12, no. 2, pp. 1731-1742, Jun. 2018.

[9]. A. Hussain, C. Xu, and M. Ali, "Security of Cloud Storage System using Various Cryptographic Techniques," International Journal of Mathematics Trends and Technology (IJMTT), vol. 60, no. 1, pp. 45–51, 2018.

[10]. L. S. Nishad, J. Paliwal, R. Pandey, S. Beniwal and S. Kumar, "Security privacy issues and challenges in cloud computing: A survey", Proc. 2nd Int. Conf. Inf. Commun. Technol. Competitive Strategies, pp. 47, 2016.

[11]. Namasudra, S., Roy, P.: Secure and efficient data access control in cloud computing environment: a survey. J. Multiagent Grid Syst. 12, 69–90 (2016)

[12]. Balogh, Z., Turčáni, M.: Modeling of data security in cloud computing. In: IEEE Annual Systems Conference, pp. 1–6. IEEE (2016)

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Masked Face Recognition and Body Temperature Detection

Sanika Aier[1], Ankita Salunke[1], Pooja Sharma[1], Sonam Patil[1], Prof.(Dr.) Pankaj Agarkar[2], Prof. Pooja Shinde[3]

[1]Student, Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Savitribai Phule Pune University, Pune, Maharashtra, India

[2]Head of Department, Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Savitribai Phule Pune University, Pune, Maharashtra, India

[3]Assistant Professor, Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Savitribai Phule Pune University, Pune, Maharashtra, India

ABSTRACT

So as to forestall the spread of CORONA otherwise known as COVID-19 infection, nearly everybody wears a veil during COVID-19 scourge. This makes the old facial acknowledgment framework ineffectual by and large, for example, network access control, face access control, facial participation, facial security checks at train stations, and so on Along these lines, it is exceptionally earnest to improve the acknowledgment execution of the current face acknowledgment innovation on the veiled appearances with internal heat level identification. Current progressed facial acknowledgment frameworks are planned dependent on profound realizing, which rely upon a more noteworthy number of face tests. Be that as it may, as of now, there are no covered face acknowledgment datasets. To this end, there are three kinds of concealed face datasets, including Masked Face Detection Dataset (MFDD), Real-world Masked Face Recognition Dataset (RMFRD) and Simulated Masked Face Recognition Dataset (SMFRD). These datasets are effectively accessible, in light of which different applications on veiled countenances can be created. So, we reason a dependable technique dependent on dispose of veiled locale and profound learning-based highlights so as to address the issue of concealed face acknowledgment measure with internal heat level identification.

I. INTRODUCTION

The COVID-19 infection can be spread through contact and surface contact, thusly, the biometric frameworks dependent on passwords or fingerprints can't be utilized further in the perspective on wellbeing. It is demonstrated that most germs are spread with our hands. In this way a contactless validation framework adequately lessens the danger of spread of disease. Face acknowledgment are more protected with no compelling reason to contact any gadget. Late investigations on COVID-19 has demonstrated that wearing a face veil by solid and contaminated populace diminishes significantly the transmission of this infection. Nonetheless, wearing the veil face causes the accompanying issues: network access control and face confirmation are become

exceptionally troublesome errands when a fabulous piece of the face is covered up by a veil. existing face acknowledgment techniques are not productive when wearing a cover which can't give the entire face picture to portrayal. uncovering the nose area is significant in the undertaking of face acknowledgment since it is utilized for face standardization, present remedy, and face coordinating. Because of these issues, face veils have essentially tested existing face acknowledgment techniques. To handle these issues, we separate two unique assignments to be specific:

1) face cover acknowledgment
2) masked face acknowledgment
3) Body temperature recognition.

The first checks whether the individual is wearing a cover or no. This can be applied in broad daylight places where the cover is necessary. Veiled face acknowledgment, then again, intends to perceive a face with cover, in light of the eyes and the brow areas. Furthermore, the third one the internal heat level as fever is one of the side effects of the COVID19 So it will check internal heat level except if it comes to (37°c) or higher. In this paper we handle the subsequent errand utilizing profound learning-based technique model so as to separate highlights from the exposed face locales (out of the veil area). It is Touchless passage framework with face/cover and fever acknowledgment. The validation framework can be incredibly useful to encourage general society in packed places, for example, clinics, schools, IT parks, and so on just as in lessening the spread of contamination.

II. RELATED WORKS:

Later designs, for example, Res Net have presented skip associations which permits further organizations to keep away from immersion in preparing precision. These structures are frequently utilized for starting element extraction in face identification organizations. In our strategy, we are utilizing VGG 16 design as the base organization for face recognition and Fully Convolutional Network for division. Despite the fact that dominant part of division designs depends on down inspecting and sequential up examining of info picture, Fully Convolutional Networks are humble and have exact methodology.

Impediment work is a critical restriction of true 2D face acknowledgment strategy. By and large, it comes out from wearing caps, eyeglasses, covers just as whatever other articles that can impede a piece of the face while leaving others unaffected. In this manner, wearing a veil is considered as the most difficult facial impediment challenge since it blocks a fabulous portion of the face including the nose. Numerous methodologies have been proposed to deal with this issue. We can group them into three classifications in particular: nearby coordinating methodology, rebuilding approach and dispose of impediment-based methodology. Coordinating methodology: Aims to analyse the similitude between pictures utilizing a coordinating cycle. By and large, the face picture is examined into various patches of a similar size. Highlight extraction is then applied on each fix. At long last, coordinating cycle is applied among test and exhibition faces. The benefit of this methodology is that the examined patches are not covered, which maintains a strategic distance from the influence of impeded locales on the other useful parts. For instance, Martinez et al eix. tested the face area into a fixed number of neighbourhood patches. coordinating is then applied for comparability measure.

Different techniques identify the central issues from the face picture, rather than neighbourhood patches. For example, Weng et al. proposed to perceive people of interest from their halfway faces. To achieve this assignment, they firstly recognized central issues and concentrate their textural and mathematical

highlights. Next, point set coordinating is completed to coordinate the acquired highlights. At long last, the comparability of two appearances is acquired through the separation between these two adjusted capabilities. Central issue-based coordinating technique is presented in Duan et al. Filter key point descriptor is applied to choose the suitable central issues. Gabor ternary example and point set coordinating are then applied to coordinate the neighborhood key focuses for fractional face acknowledgment. Rather than the previously mentioned strategies dependent on fixed-size patches coordinating or key focuses recognition, McLaughlin et al. applied a biggest coordinating zone at each purpose of the face picture with no inspecting. Reclamation approach: Here, the blocked locales in the test faces are reestablished by the exhibition ones. For example, Bag chi et al. proposed to reestablish facial impediments. The location of the blocked districts is done by thresholding the profundity map estimations of the 3D picture. At that point the reclamation is taken on by Principal Component Analysis (PCA). There are likewise a few methodologies that depend on the assessment of the blocked parts. Deira et al. applied a factual shape model to foresee and reestablish the halfway facial bends. Iterative nearest point (ICP) calculation has been utilized to eliminate impeded locales in. The reclamation is applied utilizing a bend, which utilizes factual assessment of the bends to deal with the impeded parts. Halfway watched bends are finished by utilizing the bends model created through the PCA strategy. Dispose of impediment-based methodology: In request to dodge a terrible reproduction measure, these methodologies plan to distinguish districts discovered to be blocked in the face picture, and dispose of them totally from the component extraction and classification measure. Division based methodology is perhaps the best strategy that distinguish firstly the blocked district part, and utilizing just the non-impeded part in the accompanying advances. For example, Priya and Banu separated the face picture into little neighborhood patches. Next, to dispose of the blocked area, they applied the help vector machine classifier to identify them. At last, Mean put together weight lattice is utilized with respect to the non-impeded areas for face acknowledgment. Alyuz et al. applied an impediment evacuation and reclamation. They utilized the worldwide concealed projection to eliminate the impeded areas. Next, the fractional Gappy PCA is applied for the reclamation utilizing eigenvectors. Also, Yu et al. completed a halfway coordinating system to successfully takes out the impeded areas and afterward utilizing the non-blocked locales in the coordinating cycle. Since the distribution of Alex Net engineering in 2012 by krizhevsky et al, profound CNN have become a typical methodology in face acknowledgment. It has likewise been effectively utilized in face acknowledgment under impediment variety. We find profound learning-put together technique based with respect to the way that human visual framework naturally overlooks the impeded districts and just spotlights on the non-blocked ones. For instance, Song et al. proposed a veil learning strategy so as to dispose of the element components of the concealed locale for the acknowledgment cycle. Enlivened by the superior of CNN based strategies that have solid vigor to enlightenment, outward appearance and facial impediment transforms, we propose in this paper a dispose of impediment-based strategy and profound CNN based model to address the issue of concealed face acknowledgment during COVID-19 pandemic. Exploratory outcomes are completed on Real-world Masked Face Recognition Dataset (RMFRD) introduced.

- RMFRD: A python crawler instrument is utilized to creep the front-face pictures of people of note and their comparing veiled face pictures from gigantic Internet assets. At that point, we

physically eliminate the preposterous face picture coming about because of wrong correspondence. The way toward separating pictures takes a great deal of labor. Additionally, we crop the precise face zones with the assistance of self-loader comment apparatuses, such as Labeling and LabelMe. The dataset incorporates 5,000 pictures of 525 individuals wearing veils, and 90,000 pictures of similar 525 subjects without covers. Apparently, this is as of now the world's biggest true covered face dataset. Fig. 1 shows sets of facial picture tests.

- SMFRD: In request to grow the volume and variety of the veiled face acknowledgment dataset, we in the interim have taken elective methods, which is to put on covers on the current public huge scope face datasets. To improve information control effectiveness, we have built up a cover wearing programming dependent on D-lib library to perform veil wearing consequently. This product is then used to wear veils on face pictures in the mainstream face acknowledgment datasets, as of now including LFW and Web face datasets. Along these lines, we moreover built a reenacted concealed face dataset covering 500,000 face pictures of 10,000 subjects. By and by the reenacted concealed face datasets can be utilized alongside their unique exposed partners.

III. MOTIVATION

We start by confining the cover area. Our principle reason for existing is to unhide the covered district. Right off the bat, we will begin with concealed district. To do so we apply an editing channel so as to get just the useful districts of the covered face. (for example, temple and eyes ones). By applying an editing channel to temple and eyes which isn't covered this assist with getting the basic view to coordinate the specific character.

Next, we portray the chose areas utilizing profound learning model. Profound learning model are constructed utilizing neural organizations. A neural organization takes in inputs which are then handled in concealed layers. This technique is more appropriate in certifiable applications contrasting with reclamation draws near. Notwithstanding the ongoing achievements of profound learning structures in design acknowledgment assignments they have to gauge a large number of boundaries in the completely associated layers that require amazing equipment with high preparing limit and memory.

To address this issue, we present in this paper an effective quantization based pooling technique for face acknowledgment utilizing VGG-16 pre prepared model. To do so we just consider the component maps at the last convolutional layer. The fundamental plan to speak to pictures as request less arrangements of neighborhood highlights.

SBTL8033 is a Touchless Indoor Speed Gate based passage control arrangement with internal heat level recognition capacities. The passageway control framework is incorporated with cutting edge ZK facial acknowledgment terminals highlighted with PC vision innovation and keen facial acknowledgment calculations.

This gadget joins obvious light facial acknowledgment innovation with infrared temperature recognition for exact and quick internal heat level estimation during client check.

The profound quantization procedure presents numerous preferences. It guarantees a lightweight portrayal that makes true concealed face acknowledgment an achievable errand. In addition, the covered districts change from face to another, which prompts educational pictures from various sizes. The proposed profound quantization permits grouping pictures from various sizes so as to deal with this issue. Moreover, the Deep BoF approach utilizes a differentiable quantization plot that empowers

concurrent preparing of both the quantizer and the remainder of the organization, rather than utilizing fixed quantization just to limit the model size. It merits expressing that our proposed strategy doesn't have to learn on the mission area subsequent to eliminating the veil. It rather improves the speculation of face acknowledgment measure within the sight of the cover during the pandemic of Covid.

IV. LITERATURE REVIEW

Face acknowledgment has gotten huge consideration. Yet at the same time machine acknowledgment framework has arrived at just a specific degree of development. Model, acknowledgment of face at outside climate with consistent change in stance and light is an unsolved issue.

This paper gives basic study of still face acknowledgment research. Presently, because of current COVID-19 pandemic face acknowledgment has reached to another level. Significant pieces of face i.e lips, nose, jaw are been stowing away because of face veil. Face acknowledgment simply through eyes and forehead is a major test.

Indeed, even security is additionally one of the significant factors. The solid requirement for easy to understand frameworks that can make sure about our personality in an ocean of numbers is self-evident. Albeit, entirely dependable technique for thumb print biometric individual distinguishing proof exists, it isn't protected in the present state of COVID-19 pandemic. It is demonstrated hazardous. Consequently, face acknowledgment is more secure.

V. CONCLUSION

In genuine situations (for example unconstrained conditions), human countenances may be blocked by different articles, for example, facial cover. This makes the face acknowledgment measure an exceptionally testing task. The proposed strategy accomplished a high acknowledgment execution. For the best of our insight, this is the main work that tends to the issue of concealed face acknowledgment during COVID-19 pandemic. It merits expressing that this examination isn't restricted to this pandemic period since many individuals are mindful continually, they deal with their wellbeing and wear veils to ensure themselves against contamination and to lessen different microbes' transmission.

VI. REFERENCES

[1]. Zhongyuan Wang, Guangcheng Wang, Baojin Huang, Zhangyang Xiong, Qi Hong, Hao Wu, Peng Yi, Kui Jiang, Nanxi Wang, Yingjiao Pei, et al. Masked face recognition dataset and application. arXiv preprint arXiv:2003.09093, 2020.

[2]. Nizam Ud Din, Kamran Javed, Seho Bae, and Juneho Yi. A novel gan-based network for unmasking of masked face. IEEE Access, 8:44276–44287, 2020

[3]. Sehla Loussaief and Afef Abdelkrim. Deep learning vs. bag of features in machine learning for image classification. In 2018 International Conference on Advanced Systems and Electric Technologies (IC_ASET), pages 6–10. IEEE, 2018.

[4]. Erik Hjelmås Department of Informatics, University of Oslo, P.O. Box 1080, Blindern, Oslo, N-0316 and Boon Kee Low Department of Meteorology, University of Edinburgh, JCMB, Kings Buildings, Mayfield Road, Edinburgh, Scotland, EH9 3JZ, United , Computer Vision and Image Understanding, September 2001

[5]. S Mavaddati - Journal of AI and Data Mining, 2019 - jad.shahroodut.ac.ir

[6]. Nikolaos Passalis and Anastasios Tefas. Learning bag-of-features pooling for deep convolutional neural networks.In Proceedings of the IEEE international conference on computer vision, pages 5755–5763, 2017.
[7]. Davis E King. Dlib-ml: A machine learning toolkit. The Journal of Machine Learning Research, 10:1755–1758,2009.
[8]. Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
[9]. https://www.researchgate.net/publication/263964949_Software_Hardware_for_Face_Detection
[10]. https: //arxiv.org/pdf/2003.09093.pdf
[11]. https: //eandt.theiet.org/content/articles/2020/03/facial-recognition-software-peers-behind-the-mask/
[12]. https://theconversation.com/face-masks- and-facial-recognition-will-both-be-common-in-the-future-how-will-they-co-exist-144417
[13]. https://arxiv.org/pdf/2008.11104.pdf
[14]. https://www.electronicsforu.com/new-products/ai-based-intelligent-face-mask-and-body-temperature-detection-system
[15]. https://www.embedded-computing.com/news-releases/intelligent-face-mask-and-body-temperature-detection-system
[16]. https://www.researchgate.net/figure/Masked-face-recognition-flow-chart_fig2_340690545
[17]. https://www.researchgate.net/publication/263964949_Software_Hardware_for_Face_Detection
[18]. https://www.quora.com/What-are-the-requirements-need-to-be-considered-for-face-recognition-software-i-e-software-hardware-requirements.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Flight Ticket Price Prediction Using Machine Learning

Komal Kalane[1], Shivam Ghorpade[1], Omkar Jawale[1], Abhishek Jaiswal[1], Prof. Monika Dangore[1]

[1]Computer Department, Savitribai Phule Pune University, D.Y. Patil School of Engineering Charholi, Pune, Maharashtra, India

ABSTRACT

Nowadays, airline ticket prices can vary dynamically and significantly for the same flight, even for nearby seats within the same cabin. Customers are seeking to get the lowest price while airlines are trying to keep their overall revenue as high as possible and maximize their profit. Airlines use various kinds of computational techniques to increase their revenue such as demand prediction and price discrimination. From the customer side, two kinds of models are proposed by different researchers to save money for customers: models that predict the optimal time to buy a ticket and models that predict the minimum ticket price. In this paper, we present a review of customer side and airlines side prediction models. Our review analysis shows that models on both sides rely on limited set of features such as historical ticket price data, ticket purchase date and departure date. Features extracted from external factors such as social media data and search engine query are not considered. Therefore, we introduce and discuss the concept of using social media data for ticket/demand prediction.

Keywords : Survey, Ticket price prediction, Demand prediction, Price discrimination, Social media.

I. INTRODUCTION

The airline industry is considered as one of the most sophisticated industry in using complex pricing strategies. Nowadays, ticket prices can vary dynamically and significantly for the same flight, even for nearby seats as given in paper[119-128] and paper [28-42]. The ticket price of a specific flight can change up to 7 times a day.The Cheapest available ticket for a given data gets more or less expensive over time.This usually happens as an attempt to maximize revenue based on following things-

1.Time of purchase patterns(i.e. last minute purchases are expensive)

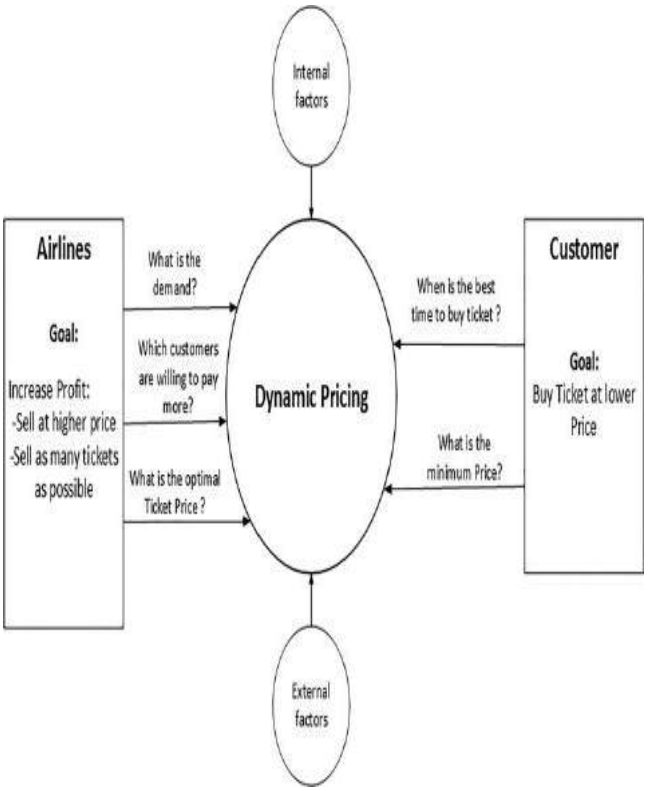
2.Keeping the flight as full as they want.

The objectives of the project can broadly be laid down by following questions:

1. Flight Trends
2. Best time to buy
3. Verifying myths

From the customer point of view, determining the minimum price or the best time to buy a ticket is the key issue. The conception of “tickets bought in advance are cheaper” is no longer working. It is possible that customers who bought a ticket earlier pay more than those who bought the same ticket later. Moreover, early purchasing implies a risk of commitment to a specific schedule that

may need to be changed usually for a fee. The ticket price may be affected by several factors thus may change continuously. To address this, various studies were conducted to support the customer in determining an optimal ticket purchase time and ticket price prediction according to papers [267-276],[10-17], ,[87-90],[84-89], [964-969].



II. RELATED WORK

In the paper[1] author proposes reports on a pilot study in the domain of airline ticket prices where we recorded over 12,000 price observations over a 41 day period. When trained on this data, Hamlet — our multi-strategy data mining algorithm — generated a predictive model that saved 341 simulated passengers $198,074 by advising them when to buy and when to postpone ticket purchases. Remarkably, a clairvoyant algorithm with complete knowledge of future prices could save at most $320,572 in our simulation, thus Hamlet's savings were 61.8% of optimal. The algorithm's savings of $198,074 represents an average savings of 23.8% for the 341 passengers for whom savings are possible. Overall, Hamlet saved 4.4% of the ticket price averaged over the entire set of 4,488 simulated passengers.
In the Paper [2] author proposes that the prices have become increasingly available on the worls wide web.They are able to comparison shop efficiently and to track prices over time.

In the paper [3] author proposes A detailed discussion on the choice of forecast models and forecast variables is reported. A salient feature of the reported methods is that the forecast models can take into account the influence of market power on electricity prices.

In the paper [4] author proposes paper presents general patterns in airline pricing behaviour and a methodology for analysing different routes and/or carriers. The purpose is to provide customers with the relevant information they need to decide the best time to purchase a ticket, striking a balance between the desire to save money and any time restraints the buyer may have.

In the paper [5] author proposes Airline ticket price depends on many different dynamic factors, such as airline pricing policies, flight distance, class of service, airline, global population mobility, all of which define the travel demand. Ticket costs can vary significantly for the same flight, even for nearby seats. The model of airline tickets market varies from country to country and depends on the volume and structure of supply (number of airlines and air flights) and demand (number of passengers, seasonal peculiarities).

In The Paper [6] author proposes Airlines are one branch of these companies, having the available seats on a plane as their standing inventory. They divide these seats into several buckets, where each bucket has its own fare price. Airlines rearrange these seats across the buckets to make more money out of them, this creates changes in the prices which customers have to pay for the flight. Such that different customers pay different prices for tickets of the same flight.

Machine Learning is an idea to learn from examples and experience, without being explicitly programmed. Instead of writing code, you feed data to the generic algorithm, and it builds logic based on the data given. The process of training an ML model involves providing an ML algorithm (that is, Machine Learning Algorithm) with training data to learn from. The data classification can be performed on structured or unstructured data. The main goal of classification is to identify the category/class to which a new data will fall under. We can use these ML models to get predictions on new data for which target is unknown. II. GENRAL STRUCTURE Fig : General Structure of ML Model The above figure represents the general structure of an ML model. The data are split into Training and Test Data respectively. The Training Data is passed to trough the ML algorithms for enabling the machine to learn and apply it on the test data to predict the solutions. The such predicted solutions can be used for comparisons, calculate the accuracy.

III. LITERATURE REVIEW SUMMARY TABLE:

Paper no.	Paper title	Methodology	Advantages	Drawbacks
1	"Mining airfare data to minimize ticket purchase price"- ETZIONIETAL (2017)	Rule learning (Ripper), Reinforcement learning (Q-learning), time series methods, and combinations of these	An average of 61.8% savings achieved as compared to optimal saving.	Limitations in data set

2	"An agent for optimizing airline ticket purchasing"- WILLIAMGRO AND MARIAGIN (2017)	PLS regression Decision tree,nu-SVRRidge Regression	75.3% saving for the as compared to optimal saving.	Does not consider heterogeneous flights
3	"APPROACH TO PRICE FORCASTING"- WOHLFARTHETAL(2018)	Marked point processes (MPP) for Preprocessing	Data collected for 28 days for 6 routes	55% performance
4	"OPTIMALPURCHASE TIMING"- DOMINGUEZ-MENCHEROET (2017)	Non-parametric isotonic regression	2 months daily price information extracted 30 days	NULL
5	"DATA DRIVEN MODELLING"- ANASTASIA (2018)	Regression Model	Ticket price data collected for 75 days	The dataset set is limited
6	"A LINEAR QUANTILE REGRESSION MODEL" -T.JANSSEN (2018)	Linear quantile mixed regression model	Performs well for shorter period	Inefficient for longer period

IV. CONCLUSION:

After surveying the papers published before we conclude that the majority of the methods used in papers above made use of traditional prediction models from the computational intelligence research field known as Machine Learning. It was difficult for the customer to purchase air ticket due to the high complexity of the pricing models applied by the airlines because the prices change dynamically. Many features that can vary the base price of a ticket was

not considered. There were problems that was encountered during manual data collection and previous works done in the field of airfare price prediction. From the experiments we concluded which features influence the airfare prediction at most. Future this project could be extended to predict the airfare prices with higher performance considering few other features.

V. REFERENCES

[1]. A. Mottini and R. Acuna-Agost, "Deep choice model using pointer networks for airline itinerary prediction," in the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, 2017

[2]. T. Liu, J. Cao, Y. Tan, and Q. Xiao, "ACER: An adaptive context-aware ensemble regression model for airfare price prediction," in the international conference on progress in informatics and computing, 2017

[3]. C. Koopmans and R. Lieshout, "Airline cost changes: To what extent are they passed through to the passenger?" Journal of Air Transport Management, vol. 53, pp. 1–11, 2016.

[4]. P. Malighetti, S. Paleari and R. Redondi, "Pricing strategies of low-cost airlines: The Ryanair case study," Journal of Air Transport Management, vol. 15, no. 4, pp. 195-203, 2009.

[5]. A regression model for predicting optimal purchase timing for airline tickets William Groves and Maria Gini October 18, 2011

[6]. Predicting Airfare Prices Manolis Papadakis, 2014

[7]. Linear Quantile Mixed Regression Model for Prediction of Airline Ticket Prices, 2014

[8]. To Buy or Not to Buy: Mining Airfare Data to Minimize, 2003

[9]. Random Forests, January 2011

[10]. Dingli, A., Mercieca, L., Spina, R., Galea, M., 2015. Event detection using social sensors. In: 2015 2nd International Conference on Information and Communication Technologies for Disaster Management (ICT-DM), Rennes, 2015, pp. 35–41.

[11]. Porshnev, A., Redkin, I., Shevchenko, A., 2013. Machine Learning in Prediction of Stock Market Indicators Based on Historical Data and Data from Twitter Sentiment Analysis. In: 2013 IEEE 13th International Conference on Data Mining Workshops, Dallas, TX, 2013, pp. 440–444.

[12]. Bo An, Haipeng Chen, Noseong Park, V.S. SubrahmanianMAP: Frequency-Based Maximization of Airline Profits based on an Ensemble Forecasting Approach Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), ACM, New York, NY, USA (2016), pp. 421-430

[13]. Bhavuk Chawla, Ms Chandandeep KaurAirfare Analysis And Prediction Using Data Mining And Machine Learning Int. J. Eng. Sci. Invention, 6 (11) (2017), pp. 10-17

[14]. Chieh-Hua Wen, Po-Hung ChenPassenger booking timing for low-cost airlines: a continuous logit approach J. Air Transport Manage., 64 (2017), pp. 91-99

[15]. Estimating dynamic demand for airlines Econ. Lett., 124 (1) (2014), pp. 26-29

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Survey on Crop Suggestion based on Regional Soil Quality

Mayuresh Kulkarni[1], Rutuja Jade[1], Apekshita Bhosale[1], Bhagyashree Ramteke[1], Dr. Sunil Rathod[2]

[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

[2]Professor, Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

Agriculture is the major source for living for the people of India and also plays a major role in economy and employment. Soil is an important key factor for agriculture .There are several soil varieties in India.. In order to predict the type of crop that can be cultivated in that particular soil type we need to understand the features and characteristics of the soil type. The common difficulty present among the Indian farmers are they don't opt for the proper crop based on their soil necessities. Because of this productivity is affected. This problem of the farmers has been solved through precision agriculture. Machine learning techniques provide a flexible way in this case. Classifying the soil according to the soil nutrients is much beneficial for the farmers to predict which crop can be cultivated in a particular soil type. Data mining and machine learning are still an emerging technique in the field of agriculture and horticulture.

KEYWORDS: Agriculture, Machine learning, Soil, Classification

I. INTRODUCTION

India is one of the biggest producers of agricultural products and still has very less farm productivity. Productivity needs to be increased so that farmers can get more pay from the same piece of land with less labor. Many researches are being carried out, in order to attain an accurate and efficient model for crop prediction.

Data mining is a challenging technology in the field of agriculture. Nowadays data mining has been used in the field of agriculture for soil classification, wasteland management, and crop and pest management [1]. In assessed the association rules of affiliation methods in DM and applied into the soil science to anticipate the significant connections and gave association rules to different soil types in agriculture. The agriculture factors such as rain, weather, soil type, pesticides and fertilizers are the main responsible to increase the production. Crop cultivation depends on the nature and the nutrients of the soil increasing the cultivation of land which brings a loss of supplements present in the soil.

Machine learning is a field of computer science where new developments evolve at recent times, and also helps in automating the evaluation and processing done by mankind, thus by reducing the burden on human power. In simple terms, machine learning

provides basic algorithms that can provide information about a dataset without writing code to solve a program manually. Instead of writing code you provide data or the basic algorithm and it forms its own conclusions based on this data. In machine learning, the methods are derived from the learning process. Those methodologies need to learn through experiences to perform a particular task.

Classification is a data mining technique based on machine learning which is used to categorize the data item in a dataset into a set of predefined classes. It helps in finding the diversity between the objects and concepts. Among these various machine learning techniques that are being used in this field; this system builds an efficient and accurate model to classify the soil type according to nutrients of soil and predict the type of crop that can be yielded in that particular soil. Also this information can further be used by the Minister of Agriculture, Food & Marketing Industrialist, etc. to grow the business and economy of the country.

II. LITERATURE SURVEY

The requirements and planning is needed for developing software model. The author's [2] start from the basics of precision farming and move towards developing a model that would support it. It deeply studies the basics of precision farming. A model that applies Precision Agriculture (PA) principles to small, open farms at the individual farmer and crop level, to affect a degree of control over variability. This model has been designed for the scenario in Kerala State where the average holding size is much lower than most of India. Hence this model can be positioned elsewhere in India only with some modifications. The comprehensive objective of the model is to deliver direct advisory services to even the smallest farmer at the level of his/her smallest plot of crop, using the most accessible technologies such as SMS and email.

The algorithms [3] used for yield prediction in this system are Support Vector Machine and K-nearest neighbour algorithm. The importance of crop selection and the factors deciding the crop selection like production rate, market price and government policies are discussed. This system will propose a Crop Selection Method (CSM) [4] which solves the crop selection problem and improves net yield rate of the crop. It suggests a series of crop to be selected over a season considering factors like weather, soil type, water density, crop type. The predicted value of influential parameters determines the accuracy of CSM. Hence there is a need to include a prediction method with improved accuracy and performance.

The system [5] aims to solve the crucial problem of selecting the classifiers for the ensemble learning. The proposal aims to achieve higher accuracy and performance. This project emphasizes the need for accuracy because it depends on the dataset and the learning algorithm. The performance of the models was compared with accuracy and computational time. It proposes a framework which would predict the production, and import for that particular year. At the end of the process we would be able to visualize the amount of production import, need and availability. Therefore it would help to make decisions on whether food has to be further imported or not. The soil dataset are analyzed and category predicted.

The future work stated is to create efficient models using various classification techniques such as support vector machine, principal component analysis soil. For crop growth to their possible, the soil must provide acceptable environment for it. Soil is the anchor of the roots. Texture determines how porous the soil is and the comfort of air and water movement

which is essential to prevent the plants from becoming waterlogged. The water holding capacity determines the crop's ability to absorb nutrients and other nutrients that are changed into ions, which is the form that the plant can use. The activity of microorganisms present in the soil and also the level of exchangeable aluminum can be affected by PH. The water holding and drainage determine the infiltration of roots. Hence for the following reasons the above stated parameters are considered for choosing a crop.

III. CRITICAL REVIEW OF RESEARCH PAPERS

Sr. No.	Title	Author	Methodology Used	Result
1.	Crop Recommendation System for Precision Agriculture	S.Pudumalar, E.Ramanujam, 2016	1. Random tree 2. CHAID 3. KNN 4. Naïve Bayes 5. WEKA tool	1. Pre-processing of data 2. Handling missing and out-of-range values 3. Feature extraction 4. Ensemble model to get higher accuracy 5. Rule generation
2.	Agriculture decision support system using data mining	Prof. Rakesh Shirsath, 2017	1.Subscription based system 2. ANN 3. Android application 4. Personalized content	1. Android app with a login module 2. Previously planted crops known to system 3. User feedback mechanism 4. Maintenance of crops.
3.	A Study on Various Data Mining Techniques for Crop Yield Prediction	Yogesh Gandge, Sandhya, 2017	1. Attribute selection 2. Multiple Linear Regression 3. Decision Tree using ID3 4. SVM 5. Neural Networks 6. C4.5 8. K-means and KNN	1. Selection of agricultural field 2. Selection of crop previously planted 3. Input from user 4. Preprocess 5. Attribute Selection
4.	RSF: A Recommendation System for Farmers	Miftahul Jannat Mukarrama, 2017	1.Location Detection 2.Data analysis and storage 3.Similar location detection 4. Recommendation generation module.	1. Physiographic, thermal, crop growing period, crop cropion rate 2. Seasonal crop database 3. Similar location detection 4. Generating the set of crops 5.Similarity between the crops planted in a region

IV. TAXONOMY CHART

Parameters → Systems ↓	K-nearest neighbour	Support Vector Machine	Internet Connection Required	Precision Agriculture	Agro-Consultant
Crop Prediction System using ML	Yes	No	Yes	No	No
Crop Recommendation System for Precision Agriculture	Yes	Yes	Yes	Yes	No
Proposed System	Yes	Yes	Yes	Yes	Yes

V. CONCLUSION

Our work would help farmers to increase productivity in agriculture, prevent soil degradation in cultivated land, and reduce the chemical use in crop production and efficient use of water resources. A model is proposed for predicting the soil type and suggests a suitable crop that can be cultivated in the soil. Our future work is aimed at an improved data set with a large number of attributes and also suitable fertilizers are suggested for the well growth of the crop cultivated.

VI. ACKNOWLEDGEMENT

It gives us great pleasure in presenting the paper on "**Survey on Crop Suggestion based on Regional Soil Quality**". We would like to take this opportunity to thank Dr. Pankaj Agarkar, Head of Computer Engineering Department, DYPSOE, Pune for giving us all the help and support we need during the course of the Paper writing work. We are really grateful to him for giving us an opportunity to work with the R&D cell of our department. Our special thanks to **Dr. Ashok Kasnale**, **Principal** DYPSOE who motivated us and created a healthy environment for us to learn in the best possible way. We also thank all the staff members of our college for their support and guidance.

VII. REFERENCES

[1]. V. Rajeshwari and K. Arunesh, " Analyzing Soil Data using Data Mining Classification techniques," Vol 9(19),May 2016.

[2]. Satish Babu (2013), 'A Software Model for Precision Agriculture for Small and Marginal Farmers', at the International Centre forFree and Open Source Software (ICFOSS) Trivandrum, India.

[3]. Anshal Savla, Parul Dhawan, Himtanaya Bhadada, Nivedita Israni, Alisha Mandholia , Sanya Bhardwaj (2015), 'Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture', Innovations in Information,Embedded and Communication systems (ICIIECS).

[4]. Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh (2015), 'Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique', International Conference on Smart Technologies and Management for

Computing, Communication, Controls, Energy and Materials (ICSTM).

[5]. Liying Yang (2011), 'Classifiers selection for ensemble learning based on accuracy and diversity' Published by Elsevier Ltd. Selection and/or peer-review under responsibility of CEIS].

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Survey on Text to Image Synthesis

Chaitanya Ghadling[1], Firosh Vasudevan[1], Ruchin Dhama[1], Shreya Lad[1], Dr. Sunil Rathod[2]

[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

[2]Professor, Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

One of the most difficult things for current Artificial Intelligence and Machine Learning systems to replicate is human creativity and imagination. Humans have the ability to create mental images of objects by just visualizing and having the general looks description of that particular object. In recent years with the evolution of GANs (Generative Adversarial Network) and its gaining popularity for being able to somewhat, replicate human creativity and imagination, research on generating high quality images from text description is boosted tremendously.

Through this research paper, we are trying to explore various GANs architectures to develop a model to generate plausible images of birds from detailed text descriptions with visual realism and semantic accuracy.

Keywords: GAN, AI, ML

I. INTRODUCTION

GAN (Generative Adversarial network) :

GANs consists of two components- Generator and Discriminator which are constantly in touch with each other working in tandem. The generator generates images and the discriminator then assess those images and provide feedback to generator about the correctness of the generated image in comparison with real images of the same object. The two neural networks constantly compete with each other to become more accurate in their predictions. The generator creates new images based on the feedback provided by the discriminator and the discriminator is trained by providing real images. The generator improves to fool the discriminator and the discriminator trains itself to not get fooled by the generator. The basic structure of GAN is shown in Fig. 1.

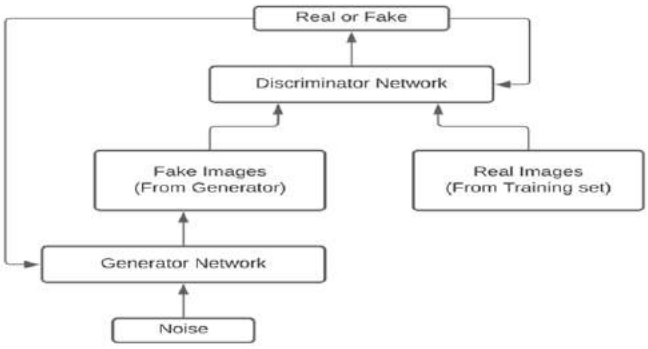


Fig. 1. Basic Structure of GAN

II. LITERATURE SURVEY

In 2014, Ian Goodfellow and his colleagues designed Generative Adversarial Network with the idea of

broadening scope of neural networks from just prediction and classification to allowing them to generate their own images. Though originally proposed as a form of generative model for unsupervised learning, GANs have also proven useful for semi-supervised learning, fully supervised learning and reinforcement learning. After various architectures developed to generate images by providing text description the quality of images along with semantic accuracy can be discussed from the Table-1.

Sr. No.	Paper Name	Advantages	Limitations
1.	Generative Adversarial Text to Image synthesis	1st major model for text to image synthesis	Lacks image quality. Does not work properly with different variety of datasets
2.	StackGAN++: Realistic Image synthesis with Stacked Generative Adversarial Networks	Improves the quality of image substantially	Difficult to train. Highly unstable and sensitive to hyper parameters.
3.	MirrorGAN: Learning Text to Image Generation by Redescription	Semantic consistency of image is highly improved.	Modules are not jointly optimized with complete end-to-end training.
4.	Learn, Imagine and Create: Text to Image Generation from prior knowledge.	Both visual realism and semantic accuracy is highly improved over baseline models.	Modules are not jointly optimized with complete end-to-end training.

Table-1 : Literature Survey

III. TAXONOMY CHART

Attributes Model	Image Quality	Semantic Accuracy	Inception Score (COCO dataset)	Inception Score (CUB dataset)
DC GAN	LOW	LOW	8.20	3.6
STACK GAN	MEDIUM	LOW	8.45	3.7
STACK GAN++	HIGH	MEDIUM	8.30	3.82
MIRROR GAN	MEDIUM	HIGH	26.47	4.56
LEICA GAN	MEDIUM	MEDIUM	20.42	4.62

Table-2: Taxonomy Chart

IV. CONCLUSION

Considering the results of the existing system, we will address the limitations on the quality of image with semantic accuracy and visual realism, by proposing a system that can improve both the parameter's and improve the overall inception score considerably. Also maintaining the consistency of output quality on other different non-standardized datasets will be one of the goals while developing a module.

V. ACKNOWLEDGEMENT

It gives us a great pleasure in presenting the paper on "A SURVEY ON TEXT TO IMAGE SYNTHESIS". We are really grateful to Dr Sunil Rathod for giving an opportunity to work with R&D cell of our department and providing us with necessary guidance with our project. We would like to take this opportunity to thank Dr. Pankaj Agarkar, Head of Computer Engineering Department, DYPSOE, Pune for providing us with an opportunity to present this paper. Our special thanks to Dr. Ashok Kasnale, Principal DYPSOE who motivated us and created a healthy environment for us to learn in the best possible way. We also thank all the staff members of our college for their support and guidance.

VI. REFERENCES

[1]. AttnGAN: Fine - grained Text to Image Generation with Attentional Generative Adversarial Networks.
[2]. StackGAN++: Realistic Image Synthesis with Stacked Generative Adversarial Networks.
[3]. Generative Adversarial Text to Image Synthesis.
[4]. https://en.wikipedia.org/wiki/Generative_adversarial_network
[5]. https://paperswithcode.com/task/text-to-image-generation

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

A Survey on 3D Model Generation from Images

Akash Chaudhari[1], Aditya Deo[1], Mahesh Badhe[1], Ritesh Patidar[1], Dr. Sunil Rathod[2]

[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

[2]Professor, Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

3D models are used for a variety of domains including video games, movies, architecture, illustration, engineering, and commercial advertising. We have seen significant progress in 3D model generation and reconstruction in recent years. There are multiple approaches, or method to do it. We discuss about different approaches in this paper, such as 3D Generative Adversarial Network (GAN), Differentiable Interpolation-based Renderer (DIB-R), Hierarchical Surface Prediction. We also discuss advantages, and limitations of these approaches. In the end it shows the results produced by them.

Keywords: 3D Model, GAN, DIB-R, Image, Neural Net

I. INTRODUCTION

A. What is a 3D Modeling ? [9]

3D modelling is the process of using software to create a mathematical representation of a 3-dimensional object or shape. The created object is termed a 3D model and these 3-dimensional models are employed in a spread of industries. The 3D modelling process produces a digital object capable of being fully animated, making it an important process for character animation and lighting tricks. The core of a model is the mesh which is a collection of points in space. These points are mapped into a 3D grid and joined together as polygonal shapes, usually triangles or quads. Each point or vertex has its own position on the grid and by combining these points into shapes, the surface of an object is created.

B. 3D Models generation from Images

Images are 2D, and they do not have complex structure, they are just 2D matrices of pixel values. Images often contain objects from real world like a car, table, or a bird. These real-world objects are 3D, and can be represented in 3D graphics also, as we see in animated movies, and games. Creating 3D model manually is very time-consuming task, therefore automation is must. Hence, 3D Model generation from images(2D). We in this survey paper try to explain the progress being made in this area.

C. What is GAN ? [9]

A Generative Adversarial Network (GAN) has two neural net (Generator and Discriminator): The generator learns to generate plausible data. The generated instances become negative training

examples for the discriminator The discriminator learns to differentiate the generator's fake data from real data. The discriminator penalizes the generator for producing implausible results. And in the end we can use that generator alone to generate our own examples/data which never existed i.e. 3D model in this case.

II. LITERATURE SURVEY

A. 3D Reconstruction from single 2D Image [2]

This was published by Deepu R, and Murali S in 2016. It uses corners of objects in an image, and then calculate depth, and length of objects. This method requires only one image. In some images, corners detection-based approach cannot give enough depth information to reconstruct 3D model. As this is not a complete ML based approach to generate 3D models, it can fail often on real world objects.

B. Hierarchical Surface Prediction for 3D Object Reconstruction [7]

Convolutional Neural Networks have shown promising results for 3D geometry prediction. They can make predictions from very little input data such as a single-color image. A major limitation of such approaches is that they only predict a coarse resolution voxel grid, which does not capture the surface of the objects well. This proposes a general framework, called hierarchical surface prediction (HSP), which facilitates prediction of high-resolution voxel grids. The main insight is that it is sufficient to predict high resolution voxels around the predicted surfaces. The exterior and interior of the objects can be represented with coarse resolution voxels. This approach is not dependent on a specific input type. It shows results for geometry prediction from color images, depth images and shape completion from partial voxel grids. Our analysis shows that high resolution predictions are more accurate than low resolution predictions in this method.

C. 3D GAN [1]

3D-GAN applied GAN in learning latent 3D space, and it can generate 3D voxel models from the latent space by extending 2D convolution into 3D convolution. Combining 3D-GAN with WGAN-GP, 3D-IWGAN can generate high-quality 3D models with a more stable training process. Wang et al. utilized an encoder–decoder as generator of the adversarial network to address 3D shape inpainting. Then a long-term recurrent convolutional network (LRCN) was employed to refine the generated results to obtain more complete 3D models in higher resolution. Chen et al. proposed text2shape system which combined 3D generation with natural language processing. The network encoded the text, then regarded the results as a condition, and utilized WGAN to decode it into a 3D model related to input text.

D. Learning to Predict 3D Objects with an Interpolation-based Differentiable Renderer (DIB-R) [5]

Unlike other approaches, this uses a whole new way to generate 3D models. It introduces a new renderer called DIB-R. DIB-R gives control to the end of how pixel value gets calculated in the rendering process of a 3D model. It introduces a new rendering pipeline just like OpenGL, and DirectX have their own pipelines. The difference is DIB-R has differentiable pipeline, which means it open doors to Machine Learning in rendering process. A GAN trained using the method proposed in this paper can predict 3D model with lighting, and texture.

The overall finding of the above discussion is given below as Table-1

© 2020 IJSRST | Print ISSN: 2395-6011 | Online ISSN: 2395-602X | Published : 18/12/2020 - 273-277

Table -1: Critical Review of Research Papers

Sr. No.	Paper Name	Advantages	Limitations
1.	Learning to Predict 3D Objects with an Interpolation-based Differentiable Renderer	It can generate 3D model with texture, and lighting with complete control of rendering	- Too much computational power requirement - Model resolution can be improved
2.	3D Reconstruction from single 2D Image	It uses corners of objects in an image, and then calculate depth, and length of objects. It requires only one image.	In some images, corners detection-based approach cannot give enough depth information to reconstruct 3D model.
3.	Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling. (3D-GAN, 3D VAE-GAN)	ML based approach, learns latent space (space where similar points are closer) from training data. It can give good results even with less than 25 training examples per class.	The model learns latent space, but cannot find relationship between different groups of similar points in the space, which can give better results.
4.	Hierarchical Surface Prediction for 3D Object Reconstruction	- Faster than GAN which uses differential rendering. - Can generate model when single input image given	- No texture, lighting with 3D model - Model is not smooth

III. TAXONOMY CHART

Attributes → Approaches ↓	3D Model Quality	Is ML based approach	Type (Image or Model)	Resources required	Hardware Required
DIB-R	HIGH	YES	Images & Models	25-30 Models 25-30 Images	GPU
3D Reconstruction from single 2D Image	LOW	NO	Image	1 Image 1 Depth Map	Basic CPU
3D GAN, and 3D VAE GAN	MEDIUM	YES	Models & Images	25 Models 25 Images	GPU
Hierarchical Surface Prediction for 3D Object Reconstruction	HIGH	YES	Image	25-30 Models 25-30 Images	GPU

© 2020 IJSRST | Print ISSN: 2395-6011 | Online ISSN: 2395-602X | Published : 18/12/2020 - 273-277

IV. RESULTS

Results of various method or approaches are shown in below Figure-1.

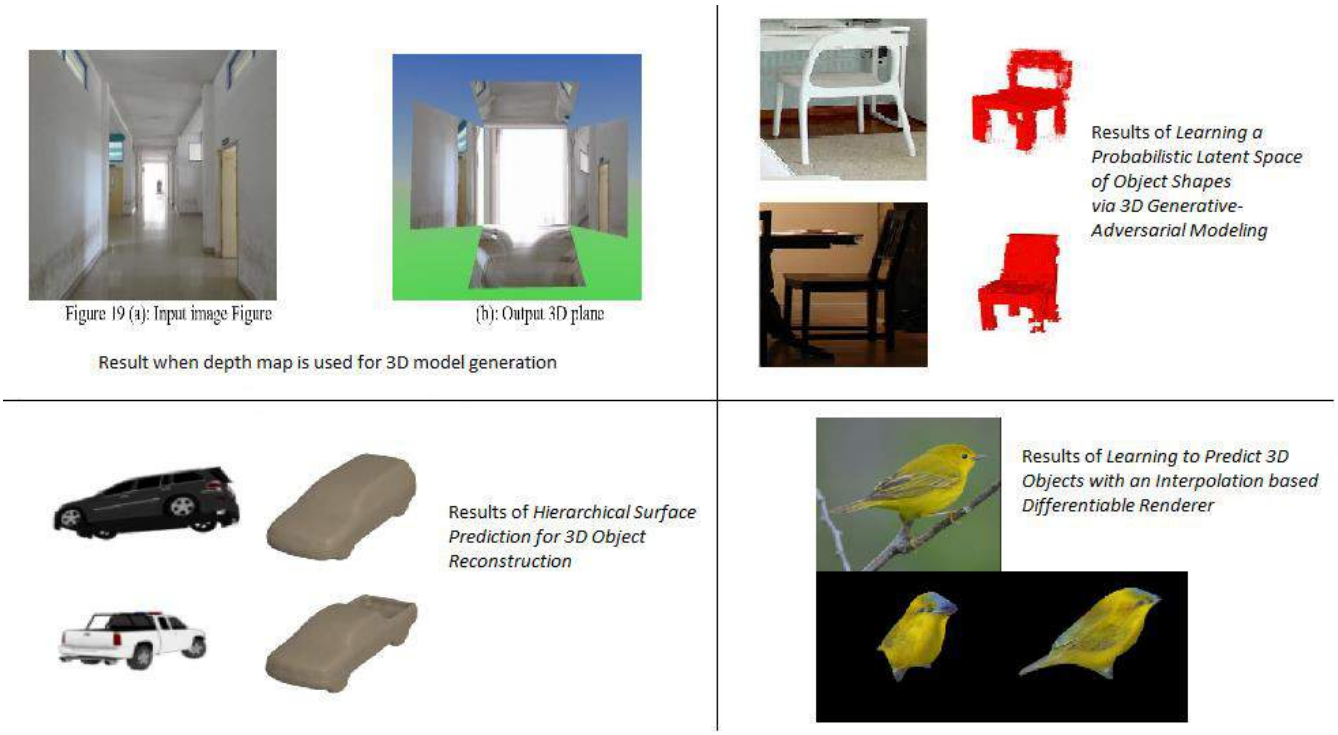


Figure-1: Results of various research papers

V. CONCLUSION

We observed that a Model should find relationship between different groups of similar points in the space, and reduce time-consumption for output generation [1]. Finding depth (z index) with a corner detection-based approach cannot give good results [2]. We also observed that if we make rendering pipeline differentiable, then it opens possibility for Machine Learning to control rendering. Taking rendering into consideration also opens possibility to predict textures, and lighting. Some approaches are very close to predict perfect 3D models which can be used in production, but still there is still need of improvement. Computational power, and data required are two main factors which need to considered.

VI. ACKNOWLEDGEMENT

It gives us a great pleasure in presenting the paper on "A SURVEY ON 3D MODEL GENERATION FROM IMAGES". We would like to take this opportunity to thank Dr. Pankaj Agarkar, Head of Computer Engineering Department, DYPSOE, Pune for giving us all the help and support we need during course of the Paper writing work. We are really grateful to him for giving an opportunity to work with R&D cell of our department. Our special thanks to Dr. Ashok Kasnale, Principal DYPSOE who motivated us and created a healthy environment for us to learn in the best possible way. We also thank all the staff members of our college for their support and guidance.

VII. REFERENCES

[1]. Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modelling, Jiajun Wu*, Chengkai Zhang*, Tianfan Xue*, William T. Freeman*, Joshua B. Tenenbaum* (2016)

[2]. 3D Reconstruction from Single 2D Image, Department of CSE, Maharaja Institute of Technology Mysore, India, January 2016

[3]. Image Generation with Disentangled 3D representation Jun Yan Zhu* (6 Dec 2018)

[4]. Unsupervised learning of 3D representations from natural images, Thu Nguyen-Phuoc*(1 Oct 2019)

[5]. Learning to predict 3D objects with an interpolation- based differentiable renderer. (2019)

[6]. https://en.wikipedia.org/wiki/3D_rendering

[7]. Hierarchical Surface Prediction for 3D Object Reconstruction Christian Häne, Shubham Tulsiani, Jitendra Malik (3 Apr 2017)

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

A Survey on Smart Digital Health Care Record with Prediction of Health Condition

Vishakha Tapkir[1], Gopal Mule[1], Aishwarya Tingre[1], Saurabh Nangare[1], Dr. Sunil Rathod[2]
[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India
[2]Professor, Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

Humans are known to be the most intelligent species on the earth and are inherently more health conscious. Since Centuries mankind has discovered various healthcare systems. To automate the process and predict diseases more correctly machine learning methods are attending popularity in research community. We need to implement machine learning methodologies to identify the best-predicted values related to the patients in their respected health condition and also need to analyze the previous health records. For that, we need to maintain a repository or the warehouse where we need to maintain digital data related to the patients and their treatment.

Keywords: Healthcare, Health Card, QR Code, Prediction, Methodology.

I INTRODUCTION

The main aim of this research paper is to store the overall health information of the patient in a Digital card. This card will consist of all the medication details, reports etc. of the patient .The implementation of the project has be done in Artificial Intelligence (Machine learning using python).

I.1 Machine Learning

Machine learning is the main background of this prediction process and the data we acquired from the medical application. This application details can be informed in the later sections and the information we gathered can be used for the machine learning models for better prediction of what is going to happen for the patient in future and what are the main constrains the patients have to follow if there are any problems with their health condition.[9][10]

I.2 Data Collection

We need to collect the data from some sources in our environment and in our circumstances. We need to create a repository which will be useful for maintaining the stability in gathering the information from other sources. In collection of information we may get noisy data and we need to remove them using pre-processing methodology. This process will maintain the stability and ambiguous information in the repository[6-9]

© 2020 IJSRST | Print ISSN: 2395-6011 | Online ISSN: 2395-602X | Published : 18/12/2020 - 278-283

I.3 Decision Tree

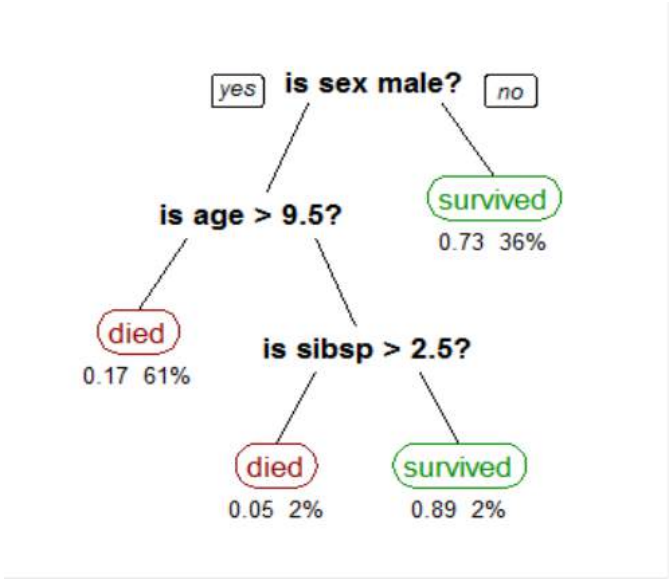


Fig-1. Decision tree example

A Decision Tree has influenced a wide area of machine learning, covering both classification and regression. In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. As the name goes, it uses a tree-like model of decisions. Though a commonly used tool in data mining for deriving a strategy to reach a particular goal, its also widely used in machine learning.[11]as shown in the fig-1.example

I.4 Random Forest Tree

Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction (see fig2. example).

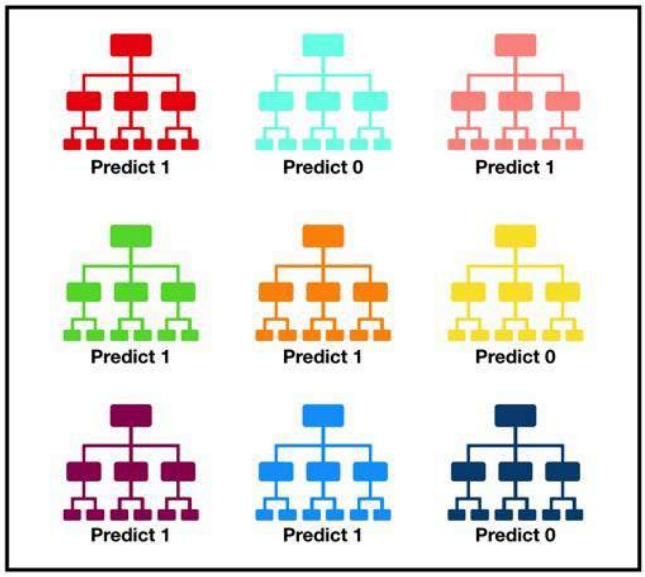


Fig-2.Random forest tree example

In data science speak, the reason that the random forest model works so well is:A large number of relatively uncorrelated models (trees) operating as a committee will outperform any of the individual constituent models.

I.5 Digital Card

The Digital Card consists of the QR Code scanner which can be only accessed by the doctor and the receptionist in the hospital. The QR Code looks like as follows [16][17].

This QR Code contains the following information related to the patient and their treatment information a. Name b. Gender c. Age d. Weight e. Height f. Contact info g. Habits h. Food Type i. Any Previous Disease information j. Duration of the suffering k. Current symptoms. l. Lab reports m. Scanning Reports n. Medication History o. Current Medication p. Current Review updatesq. Current Lab Reports (If Any) r. Next Review Updates.

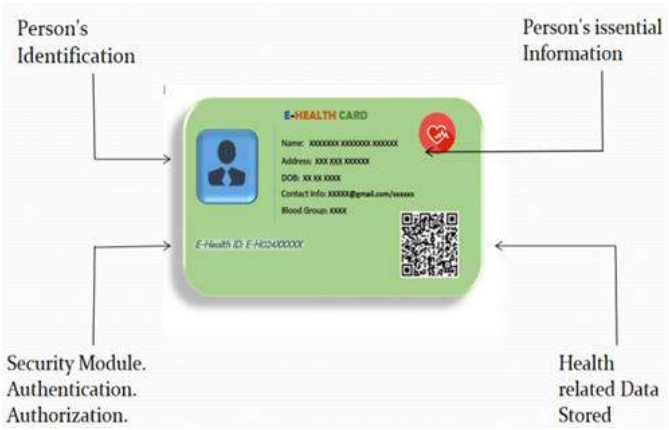


Fig-3.Layout of Health Card

It explains the structure of the medical card and the information related to the patients can be stored in the QR Code which we are generating on the card.

II LITERATURE SURVEY

Healthcare Models Smart cards offer a new perspective for healthcare applications due to the security level provided for data storage. Smart cards in healthcare applications can be used for storing information including personal data, insurance policy, emergency medical information, hospital admission data and recent medical records [3]. Numerous healthcare systems around the world start using smart cards to improve the quality of healthcare services [4]. Different healthcare models have been proposed either on national level; e.g. in Germany [5] or regional level; e.g. in US. [6].

In the field of machine learning and artificial intelligence, we can look up to several ways to predict health condition, maintain the records and get rid of Inconsistencies, redundancy and data loss. The same is achieved by:-"Disease prediction technique using data mining techniques" which pacts with the prediction of heart diseases using different machine learning algorithms. It excerpts raw data from the database and provides analysis of the working of the heart. "Computer based diagnosis using artificial neural network"

II.1 Research

It consists of the information related to different research identifications in the field of medical. NCBI is the main repository for the medical information identification. All the medical records and the information related to different researches in the medical domain can be identified in NCBI repository.[1] Here we gathered information related to some of theresearch works over the globe. They are as follows:

II.1.1 Jougen Research

Jougen [1] and his team worked on the principle of identifying the comparison between different paper based and electronic health records of the patients. EHR also referred as Electronic Health Records is the base function of many other things like digital maintenance of the reports, files, care sheets etc. All the information which are on the paper can be converted to the digital records using EHR. EHR is a very powerful thought of mankind to maintain health records into digital format.

II.1.2 Kathrin M Cresswell

Kathrin M. Cresswell [3] Deals with the format of EHR maintenance in the health record maintenance in this research work mentioned in the science direct. He was dealing with the inpatient clinical records maintenance of different hospitals.

II.2 Existing System

There are many different designed on the health care most effective thing in the medical there will be a large amount manipulated and the performance machine learning models will best features we include in our systems of this kind of work eQR Code and data security The QR Code will help the doctor best way to treat the patient information.

The current mainly focusing on how the repository and not even process of maintaining those performance of predictions [4].

The overall finding of the above discussion is given below as Table-1.

Table-1: Comparative review of Existing Research Papers

Sr. No.	Paper Name	Author	Method Proposed	Limitations
1.	A Smart Card Based Healthcare System	Hakim Fourar-Laidi E. Turhan Tunali	Access accurate health data quickly. Encryption Keys and Digital signature. Software Reusability.	Internet supported system ,Network issue. Technical problem risks are high.
2.	A Case Study For Bangladesh for Healthcare System	Prof. Mohammad Ariful Bashed. Palash Chandra Roy	Secure and Authenticated and Data Communication. Speed ,Portability Efficient to use and easy interface.	Less Cost-efficient. User needs to put correct data or else it behaves abnormally.
3.	Electronic Healthcare Model Based on Smart Card For Saudi Medical Centres.	Ebtisam Alabdulqader Hakim Fourar-Laidi	Pharmacies provide the prescription only when insurance company allows. Synchronization system synchronizes data every time patient uses card.	Redundancy of data is seen. No unique ID is provided to the card.

III TAXONOMY CHART

Parameters→ **Papers ↓**	**QRCode On Card**	**Downloadof Health Information**	**Trackingof Data Using unique ID**	**Health Prediction**	**Easeof Access**	**Stake holder Iinvolvement**
ASmart Card Based Healthcare System	NO	NO	NO	NO	YES	YES
A Case Study For Bangladesh for Healthcare System	NO	NO	YES	NO	YES	YES
Electronic Healthcare Model Based on Smart Card For Saudi Medical Centres	YES	NO	NO	YES	NO	YES

Table-2: Taxonomy chart

IV RESULT

Decision trees and random forests are supervised learning algorithms used for both classification and regression problems. These two algorithms are best explained together because random forests are a bunch of decision trees combined.[7] There are ofcourse certain dynamics and parameters to consider when creating and combining decision trees.

Initially data will be generated application and in the first phase decision tree model and after results we transfer the remaining random forest for the better prediction Here we get the highest accuracy forest than DT.[2] Fig-4 will represent the random forest and decision the result of the disease gathered.

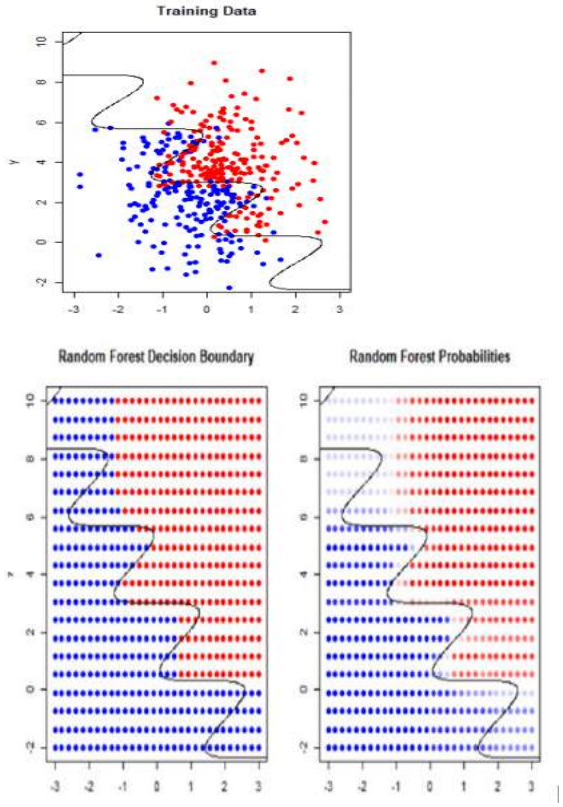


Fig-4: Result of Training data with Decisiontrees and the random forest decision boundary implementation.

V CONCLUSION

The main purpose of this paper is to determine work on medical database with the help of digital card to store information and analyse.This research paper focuses on storing the patients' health information in the digital card,analysing and designing a system where patients real-time information can be processed and evaluated based on previous symptoms and on current symptoms for different diseases. By this paper we have concluded that Random Forest, Decision tree are the best algorithms with higher accuracy rate than others for predicting and analysis.[13] So in future we can continue this paper by implementing these algorithms for better results and working model. This paper also outlines the technique to deploy this method to android and web platform to analyse and predict using real time data of users by collaborating with doctors and various medical organization.[5]

VI ACKNOWLEDGEMENT

It gives us a great pleasure in presenting the paper on "A SURVEY ONSMART DIGITAL HEALTH CARE RECORD WITH PREDICTION OF HEALTH CONDITION". We would like to take this opportunity to thank **Dr. Pankaj Agarkar**, Head of Computer Engineering Department, DYPSOE, Pune for giving us all the help and support we need during course of the Paper writing work. We are really grateful to &**Dr. Sunil Rathod** for giving an opportunity to work with R&D cell of our department. Our special thanks to **Dr.AshokKasnale,Principal DYPSOE** who motivated us and created a healthy environment for us to learn in the best possible way. We also thank all the staff members of our college for their support and guidance.

REFERENCES

[1]. M.A.NisharaBe.anu, B.Gomathy, "Disease Predicting System Using Data Mining Techniques", International Journal of Technical Research and Applications e-ISSN: 2320-8163,www.ijtra.com Volume 1, Issue 5 (NovDec 2017), PP. 41-45.

[2]. M.A.NisharaBanu, B.Gomathy, "Disease Predicting System Using Data Mining Techniques", International Journal of Technical Research and Applications e-ISSN: 2320-8163,www.ijtra.com Volume 1, Issue 5 (NovDec 2017), PP. 41-45.

[3]. "Inpatient clinical information system", Kathrin M. Cresswell, 2017, Science Direct.

[4]. https://www.researchgate.net/publication/332188767

[5]. A. Danny, S. Li, P. Houle, M. Wilcox, R. Phillips, P. Mohseni, S. Zeiger, H. Bergsten, M. Ferris, J. Diamond, M. Bogovich, M. Fleury, K. Vedati, A. Halberstadt and A. Patzer, Professional Java Server Programming: with Servlets, Java Server Pages (JSP), XML, Enterprise Java Beans (EJB), JNDI, CORBA, Jini and Java spaces (Wrox Press Inc., USA,1999).

[6]. D Kumar, R Singh, A Kumar, N Sharma An adaptive method of PCA for minimization of classification error using Naïve Bayes classifier Procedia Computer Science, 2015. Elsevier, pp.9-15.

[7]. Kumar, A., & SAIRAM, T. (2018). Machine Learning Approach for User Accounts Identification with Unwanted Information and data. International Journal of Machine Learning and Networked Collaborative Engineering, 2(03), 119-127.

[8]. Rawat K., Kumar A., Gautam A.K. (2014) Lower Bound on Naïve Bayes Classifier Accuracy in Case of Noisy Data. In: Babu B. et al. (eds) Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012), December 28-30, 2012. Advances in Intelligent Systems and Computing, vol 236. Springer, New Delhi DOI: https:// 10.1007/978-81-322-1602-5_68.

[9]. https://towardsdatascience.com/decision-trees-in-machine-learning-641b9c4e8052

[10]. I. J. Goodfellow, D.Warde-Farley, M. Mirza, A. Courville, and Y. Bengio. Maxout networks. In Proceedings of the 30th International Conference on Machine Learning, pages 1319-1327. ACM, 2013.

[11]. G. Hinton and R. Salakhutdinov. Reducing the dimensionality of data with neural networks. Science, 313(5786):504-507,

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Smart Passenger

Prashant Gaikwad[1], Akshay Mali[1], Akash Kurund[1], Manthan Khariwal[1], Varsha Babar[1]
[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

Smart Passenger is a smart device which is used to keep driver away from his drowsiness and makes his journey peaceful. Smart Passenger keeps interacting with the driver so that he stays awake during his entire journey. Smart passenger has features such as mist spraying of water, alarm system, alcohol detection, etc. Smart passenger interaction with the driver will be completely based upon driver's interest as well as his hobbies. A camera fixed with this device keeps track over the driver's movement of eyes and face using ML algorithms. Once the drowsiness is detected, stage of drowsiness will be checked & appropriate action is performed.

KEYWORDS - Machine learning, IoT, NLP, Drowsiness.

I. INTRODUCTION

Based on a study by AAA Foundation on Traffic Safety about 328,000 drowsy driving crashes occur annually. The same study found that 109,000 of those drowsy driving crashes resulted in an injury and about 6,400 were fatal. To reduce these accidents, we have come up with a smart device which acts as a companion with a driver during his entire journey and keeps him awake. This device has a built-in camera which keeps track on the eyes and mouth of the driver. It also has a microphone which is used by driver to interact with device. Based on these inputs, this device alerts the user in various ways such as mist spraying of water or raising an alarm. Its main functionality is to interact with the driver and asking him different set of questions based upon his/her interest. Driver has to fill few information in his account on the Smart Passenger website where he can upload his area of interest, hobbies, etc. all the things which he loves to talk about. Then this device uses this given data and ask question based on it. Suppose if the driver gives interest as "Cricket" then the device will ask questions like "Hey buddy, who is your favourite cricketer?" or "What do you like most about cricket?". Also, device will ask questions based on his personal information such as "Which is your favourite holiday destination?" and if the user answers wrong or answers late, then device performs appropriate actions. Apart from this, it also has a GPS inside which helps him send his location to the specified contact. Not only device asks question but also the driver can ask questions such as "Hey smarty, what's the weather condition in Pune?". Smart passenger gives real time information to the driver fetching data from the internet. Once the driver is found drowsy, first step is to keep interacting with him. If still he feels drowsy, then the next step is to mist spray of water over his face to make him feel fresh enough. This water will be in a very little amount and won't

stumble. Still if he feels drowsy, then an alarm is raised with a high volume so that he wakes up as well as nearby passing vehicles also be alert from the driver. It contains an alcohol sensor which detects if the driver is drunk. If found drunk, it asks driver not to drive car and also sends a message with location to the given contact number.

II. Functions Of Smart Passenger

A. Alcohol Detection

Alcohol detection is one of the features in this device which uses sensor to detect alcohol and raise an alarm for warning other passing cars.

MQ3 sensor is used in detecting the alcohol.

MQ-3 gas sensor has high sensitivity to Alcohol, and has good resistance to disturb of gasoline, smoke and vapour. This sensor provides an analogue resistive output based on alcohol concentration. When the alcohol gas exists, the sensor's conductivity gets higher along with the gas concentration rising. Whenever the user enters the car, the stinky smell of alcohol is detected by this sensor and informs the microprocessor to raise an alarm.

Figure.2.1 MQ3 Sensor

B. Water Spray

Water spray is key functionality of this device. This device sprinkles a very low amount of sweet smell water over user face to make him feel fresh. This spray won't stumble driver as the amount of water is very less. This uses a small refillable bottle with a spray notch head over it. It is actuated using a 100 RPM centre shaft DC motor. With the help of this motor, the spray nozzle is actuated.

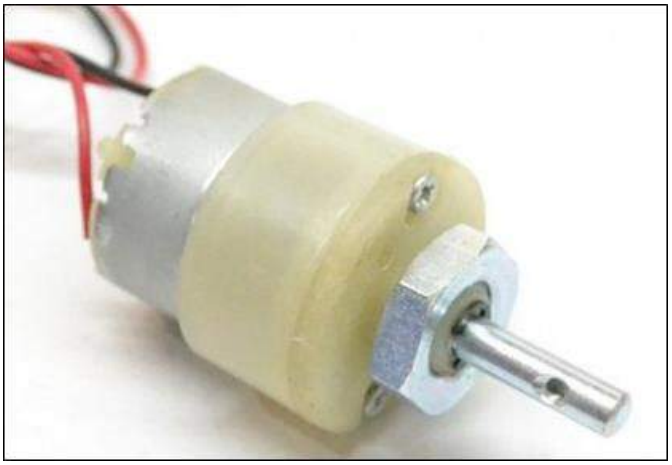


Figure. 2.2 DC Motor

C. Realtime Location

If the driver is drunken, then his real time location is sent to the specified mobile number which he has added to his profile as an emergency contact. This will help his relatives to track him if he meets with an accident due to drunken drive. Also, in future scope this location can be forwarded to nearby police stations so that cops can trace him. It uses a REB 4216 GPS module to locate the driver. Based on the latitude and longitude, the driver's location can be found. This will get internet connectivity through a GSM module which consist of only internet scheme.

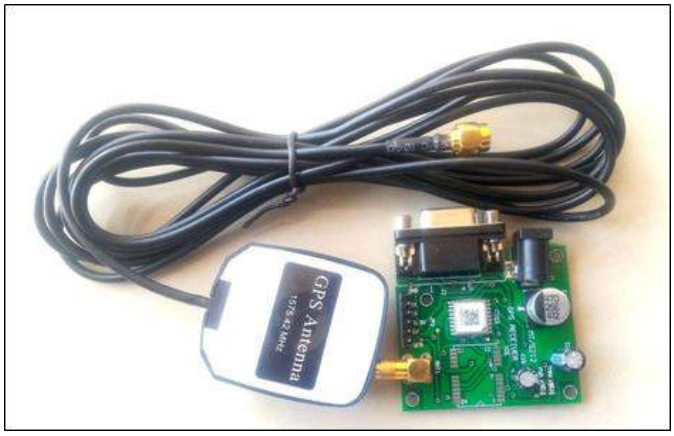


Figure.2.3 GPS Module

D. Drowsiness Detection

This device's main functionality is drowsiness detection. Using computer vision algorithms, we can automatically detect driver's drowsiness in real time video stream and then take appropriate action. It uses Raspberry Pi Camera module to detect drowsiness. To detect drowsiness, we first have to detect eye blinks in video stream. If the eye is closed for more than a specified threshold, then drowsiness is detected. This is done by finding Eye Aspect Ratio (EAR). It is represented by ∑ (x, y) co-ordinates, starting from the left corner of an eye and then working clockwise around remainder of the region.

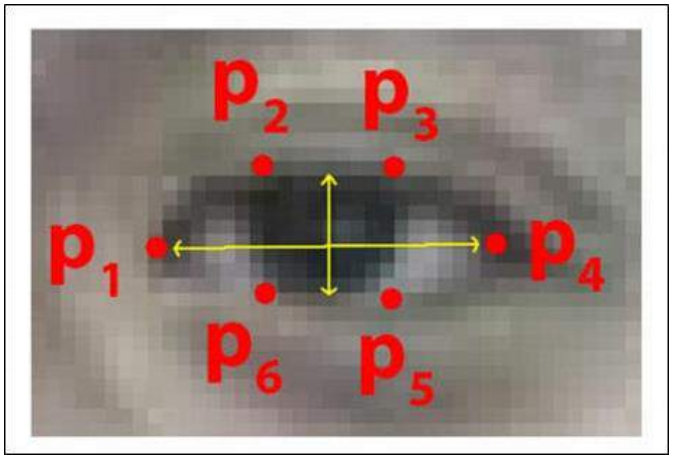


Figure.2.4 EAR

So, after getting these 6 co-ordinates of an eye, we can then derive an equation that reflects thin relation called the Eye Aspect Ratio.

$$EAR = \frac{|| P_2 - P_6 || + || P_3 - P_5 ||}{2|| P_1 - P_4 ||}$$

Where p1, p2…., p6 are facial landmark locations. The numerators of this equation compute the eye verticals landmarks, while the denominator computes the distance between horizontal eye landmarks. Since there is only one set of horizontal points but two sets of vertical points.

So, how EAR equation detects the blink of eye? Well, the EAR is approximately constant while the eye is open but will rapidly fall to zero when an eye blink occurs.

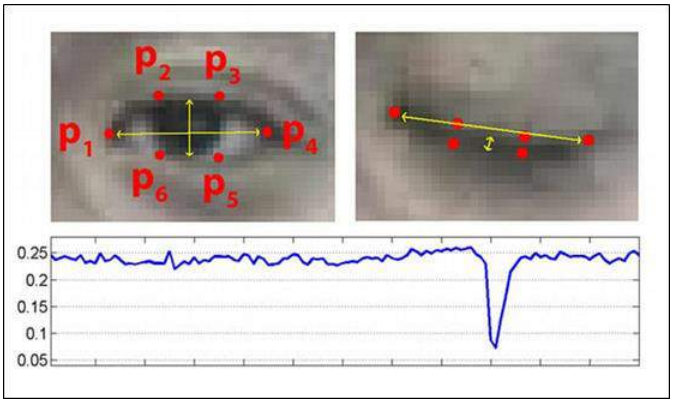


Figure.2.5 EAR Calculation

So, consider Fig.2.5, on the left side we have an eye which is fully open – the EAR is large and constant over time. However, once the person blinks the eye, EAR decreases to zero, then rapidly increases again, indicating a single blink has taken.

Using same technique, we compute eye aspect ratio for each eye, which gives us singular value. We than have declare the constant "Threshold" i.e. 0.25 & "Frame check" i.e. 20 in our algorithm.

So, if eye aspect ratio falls below this threshold, we will start counting number of frames the person has closed their eyes for. If the person's eye is closed for more than 20 frames, then it will be detected as drowsy and appropriate action will be taken.

E. Human Machine Interaction

By using tokenization algorithms in NLP (naive Bayes and LSTM), human machine interaction is made possible. Data generated from conversation are unstructured data. Unstructured Data usually do not fit properly into the traditional row and column structure of databases, and represents the vast majority of data available. Similarly, whatever user speaks through microphone is unstructured and we need to handle this which is very messy and complex part. During this process we use tokenization technique of NLP where segments of the voice are broken into small piece of text called token where each token will be a word spoken by the user from microphone. Suppose if user says "Hey Smarty, what's the weather condition in Pune?" While driver is speaking this tokenization, technique breaks it into tokens like [Hey][Smarty][what's][the] [weather][condition][in][Pune]. Once it is broken into tokens it becomes easy to understand what user said by accessing these tokens. Tokenization removes the commas, question mark, etc so that processing time is reduced.

III. Working Of Smart Passenger

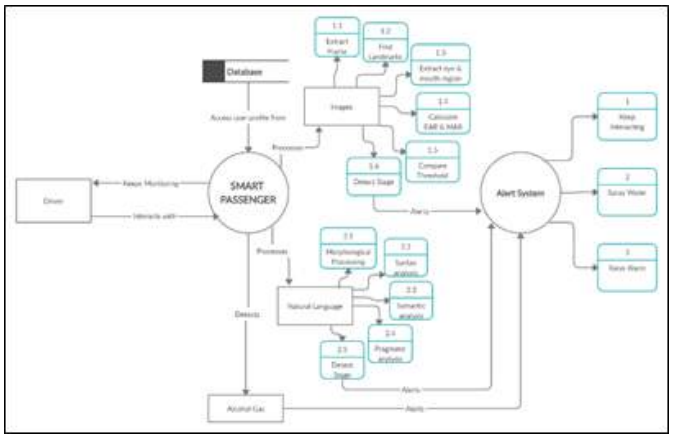


Figure. 3.1. Working of Smart Passenger

Working of Smart Passenger is very simple, driver first interacts with smart passenger using an inbuilt microphone. At the same time, this device keeps monitoring driver's face to detect drowsiness. This thing continues in a loop for long time. If any alcohol gas is detected inside the car, MQ3 Sensor will detect the gas and makes a call to alert system.

For drowsiness detection it, performs several operations such as extract frame, find facial landmarks, calculate EAR, etc. Based on that decision is made whether driver is drowsy or not. If drowsy, then a call to alert system is made.

For Natural Language Processing, the input taken from microphone is given to NLP algorithm. Then the algorithm processes the sound with different operations such as morphological processing, syntax analysis, semantic analysis, etc. Based on that smart passenger interacts with the driver. Also, if the user fails to answer the question to in given time then a call to alert system is made.

Questions asked to the driver are based on the profile which driver filled while creating an account on smart passenger official website. This data is accessed over the internet into the smart passenger and based on this the entire conversation is planned.

IV. Applications

- Smart Passenger reduces road accident cases caused due to drowsiness of driver.
- It provides a virtual travel partner when a driver is on a solo trip for a long distance.
- It can be used in cars, trucks, trains, etc.
- It reduces health risk caused due to anti sleeping pills taken by driver during long distance journey.

V. Acknowledgments

The authors would like to thank to Department of Computer Engineering of D. Y. Patil School of Engineering & Technology, Pune for its generous support. They would also like to thank the Principal Mr. Ashok Kasnale, HOD Mrs. Bhagyashree Dhakulkar and all the staff members for their valuable guidance.

VI. CONCLUSION

We identified a major problem in the society which was car accident due to drowsiness and found a solution of installing a Smart Passenger in the car so that driver doesn't fall asleep during his journey. Also, we stated many different features of Smart Passenger such as Human-Machine interaction, Alarm, Alcohol detection, etc.

VII. REFERENCES

[1]. Deep Learning for Natural Language Processing & Language Modelling by Mr. Piotr Klosowski SAP 2018, September 19th -21st, 2018, Poznan, POLAND.

[2]. A Survey on State-of-the-Art Drowsiness Detection Techniques by Muhammad Ramzan, Hikmat Ullah Khan, Shahid Mahmood Awan, Amina Ismail, Mahwish Ilyas and Ahsan Mahmood in IEEE Access on 1st May 2019.

[3]. Telematics: Artificial passenger & beyond by Mr. Dimitri Kanevsky at IBM, T. J. Watson Research Center.

[4]. Drunken driving detection and prevention models using Internet of Things by Suparna Sahabiswas and Sourav Saha.

[5]. Artificial intelligence techniques for driving safety by Zahid Halim, Rizwana Kalsoom, Shariq Bashir and Ghulam Abbas at Artif Intell Rev DOI 10.1007/s10462-016-9467-9

[6]. Facial expression recognition using face-regions by Khadija Lekdioui, Yassine Ruichek, Rochdi Messoussi, Youness Chaabi and Raja Touahni in 3rd International Conference on Advanced Technologies for Signal and Image Processing - ATSIP'2017, May 22-24, 2017, Fez, Morroco.

[7]. Research on Speech Recognition Technology and Its Application by Youhao Yu in 2012 International Conference on Computer Science and Electronics Engineering.

[8]. Driver fatigue detection system by Yogesh Chellappa, Narendra Nath Joshi, and Vaishnavi Bharadwaj in 2016 IEEE International Conference on Signal and Image Processing.

[9]. Real-Time Eye Blink Detection using Facial Landmarks by Tereza Soukupova and Jan Cech in 21st Computer Vision Winter Workshop Luka Cehovin, Rok Mandeljc, Vitomir Struc (eds.) Rimske Toplice, Slovenia, February 3–5, 2016.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Development of Tumorhunt Algorithm for Brain Tumor Segmentation Using Machine Learning CNN Model

Smita Kakade[1], Dr. Vinod M. Vaze[2], Dr. Pankaj Agarkar[3], Deepali Hirolikar[4]

[1]Research Scholar Computer Department JJTU, Jhunjhunu, Rajasthan, India

[2] Rtd. MCA Coordinator., MU, Professor Computer Sc., JJTU, Jhunjhunu, Rajasthan, , India

[3]D.Y.Patil COE, Lohagaon, Pune 411032, Maharashtra, India

[4]Research Scholar Computer Department JJTU, Jhunjhunu, Rajasthan, India

ABSTRACT

Detection of brain tumor requires brain picture segmentation as well as, manual discovering brain MR pictures segmentation is a hard task. It needs the required time, non-recurring activity, non-uniform Segmentation and in addition segmentation outcomes can vary greatly from professional to expert. The first discovery of brain tumor is essential to lessen the mortality price in patients. There is a very urgent need of automatic brain image segmentation. Hence, proposed work focused on development of machine learning algorithm for automatic segmentation. This paper presents the latest literature review and newly developed "TumorHunt" algorithm

Keywords: Image Processing, Machine Learning, CNN, Segmentation

I. INTRODUCTION

There are numerous benefits of applying virtual images approaches just like data isn't changed when it's produced over and over and preserves ingenuity, improvement of pictures causes function less difficult for doctors to understand as well as , speedy assessment of images [1,2,3]. Image segmentation is a significant and so difficult element in the medical graphic segmentation. Automatic brain tumor detection [4] and segmentation [5] encounters various problems and challenges. It is a trial to segment brain tumor within an automatic digital program as it entails diagnosis, physics linked to MRI along with strength as well as, structure evaluation of MRI image. The significant concern with brain tumor segmentation is usually which usually any tumor differs in type of shape, size, picture intensities and area. Hands-on segmentation of brain tumor mandates human being specialists and so it requires lots of time, making a pc assisted program meant for brain tumor recognition as well as , segmentation an appealing approach [6]. Diverse techniques for brain tumor detection and segmentation have already been proposed. A brain tumor can be an abnormal development of body cells within just the brain, which may be malignant or perhaps noncancerous (harmless). Magnetic resonance imaging (MRI) gives comprehensive information regarding brain tumor structure as well as , functions as an important pre processing stage with respect to tumor recognition [7,8]. Appropriate measurements in brain diagnosis are very challenging due to various shapes, appearances of tumors and sizes. Tumors may develop quickly leading to disorders in border cells additionally, gives a standard unusual framework for

healthier tissues as well. MRI is actually found in the biomedical to discover and envisage better information in the inner structure of your body [9]. This technique is actually utilised to find any variations in the cells that have a much better procedure when compared with calculated tomography. Which means this creates this system an extremely particular 1 for the brain tumor recognition as well as , malignancy imaging.

To get correct image evaluation, sharpening among any image is definitely required which may be attained by applying diverse strategies. Seeing that right away sound has been eliminated through the use of several low pass filter systems [10], we have to sharpens the picture as we are in need of the razor-sharp edges considering this can help us which can identify the border of the tumor. Existing methods are available but there is a need of deep learning practices to maintain high accuracy. Hence, this paper provides details of new algorithm named "TumorHunt" which is developed using machine learning and deep learning models.

II. METHODS AND MATERIAL

Unsupervised learning centered clustering has been effectively utilized for brain tumor segmentation by group info predicated on specific likeness requirements. Author mixed fuzzy clustering because of region-growing meant for brain tumor instances scanned by T1- weighted as well as , T2-weighted sequences and accomplished a segmentation precision of 73%. A multi-level fuzzy c-means framework was proposed to portion brain tumors diagnosed by means of multimodal MRI as well as , acquired possible outcomes, however the suggested structure was examined on an extremely modest quantity of datasets [11].

Most recently, a study is actually completed to examine diverse clustering algorithms pertaining to glioblastoma segmentation, as well as , outcomes demonstrated which usually Gaussian hidden Markov random discipline perform better k-means, fuzzy k-means and Gaussian combination model because of this activity. Nevertheless, the very best executing algorithm explained in this research still just accomplished 77% precision. However, supervised learning centered strategies require teaching data as well as , label pairs to understand a classification unit, predicated on that new situations could be categorized and segmented [12].

Author exercised super pixel aspects in a conditional arbitrary areas platform to segment brain tumors, however the results varied considerably between diverse individual conditions and specifically underperformed in LGG pictures. A report was proposed where really randomized forest was utilized meant for classifying both overall look as well as , context established features nd 81% Dice rating was achieved. Recently, author even more bundled extremely randomized trees classification with very pixel based mostly over-segmentation pertaining to an individual FLAIR sequence founded MRI scan that acquired 84% general Dice rating of the entire tumor segmentation for both LGG as well as , HGG tumor instances. However, the tuning of super pixel size and simplicity could possibly be tricky as well as , impact the ultimate delineation [13].

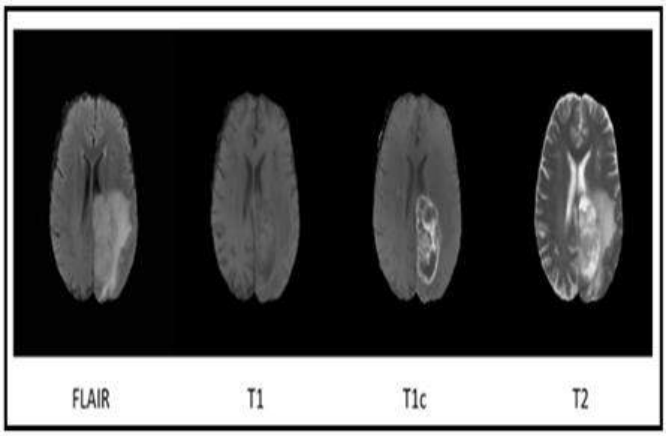


Figure 1: Stages of brain tumor

The majority of the recent conventional analysis techniques derive from human encounter in interpreting the MRI-scan for opinion; certainly this escalates the likelihood to fake detection as well as , recognition of the brain tumor [14]. However, employing digital graphic refinement guarantees the speedy and precise recognition of any tumor. Probably the most effective ways to acquire info via complicated medical images which has diverse software in medical field may be the segmentation procedure. The primary goal of the image segmentation is usually to partition a graphic into mutually unique and so exhausted areas in a way that every area of interest is usually spatially continuous as well as , the pixels within just the spot are homogeneous regarding a predetermined requirements.

Trusted homogeneity requirements consist of ideals of intensity, consistency, color, range, area curvatures and surface normal. Color centered segmentation employing K-means clustering [15] intended for brain tumor detection has were proposed, where better results had been attained using the designed algorithm than that in additional edge recognition algorithms. A altered technique was suggested that also considers the proportion evaluation and any specific vital preceding details of the region of interest along with the region as well as , edge data in the tumor area of pathological instances.

Edge depending segmentation [16] is detailed when it comes to discontinuities in picture characteristics as Gray level, texture, color etc. These types of discontinuities will be referred to as edges and so are recognized using edge recognition operators, a few of the generally utilized operators will be Prewitt, Laplace, etc. The Canny algorithm [17] may be used an optimal advantage detector predicated on a couple of specifications such as locating the most edges by reducing the mistake rate, marking edges as carefully as feasible to the real edges to increase localization, as well as , tagging edges only one time when a solitary border prevails for little results.

Lately, supervised deep convolutional neural systems (CNN) (refer figure 2 below) [18] possess captivated plenty of interests. In comparison to conventional supervised machine learning strategies, these kinds of deep learning centered methods aren't reliant on manual offers, however, instantly find out a structure among progressively intricate features straight via data.

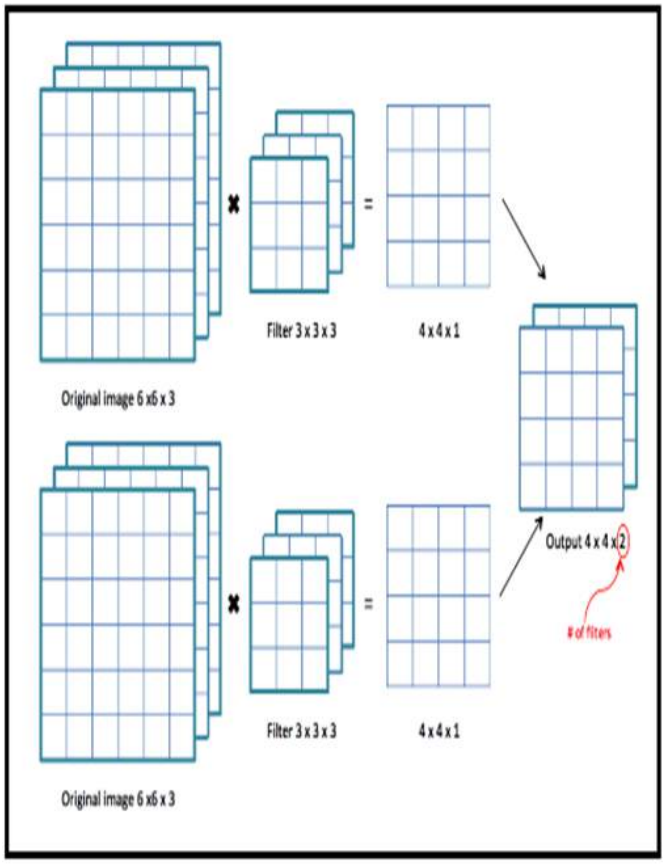


Figure 2: Representation of CNN process

Presently, using open up datasets and their particular benchmarking program, deep learning centered methods have already been scored along with the event. This is often imputed to the actual fact that deep CNN is usually built by means of putting many convolutional layers, that require convolving a sign or a graphic because of kernels to create a hierarchy of

features that will be better quality as well as , adaptable for the discriminative versions.

Medical imaging performs an essential part in disease analysis as well as , treatment preparation and also medical monitoring. The assortment of MRI accumulation concerning its configurations and geometry (2D or 3D) likewise the major difference in machine may deliver deviation in the looks of the tumors which makes the automatic segmentation complex. A precise brain lesion segmentation algorithm predicated on multi-modal MR pictures could probably enhance the conjecture precision and so effectiveness for an improved solution setting up as well as , monitoring the condition progress. CNN's have already been trusted in picture processing problems.

Convolutional networks be capable of acquire a hierarchy structure of progressively intricate aspects making them very attractive. That is done by dealing with the result attribute maps of a convolutional coating as input stations to the next convolutional layer.

One issue by way of various machine learning methods is usually which usually individuals carry out pixel category devoid of considering the neighborhood dependencies of labels. Otherwise, you can style label dependencies by taking into consideration the pixel-smart likelihood shows of a short CNN as extra insight to prepared pictures by way of TumorHunt algorithm. The detailed flowchart is shown in figure 3 below.

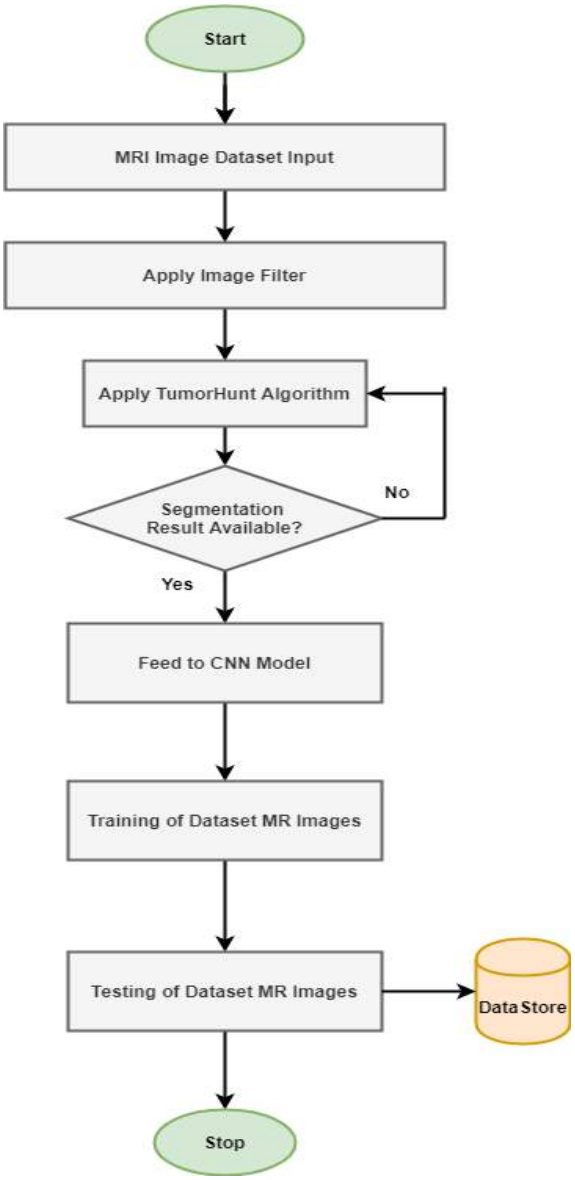


Figure 3: Proposed system flowchart

TumorHunt Algorithm:

1. Store dataset of brain MR images
2. Apply Gabor filter to reduce noise level
3. Array imgArr [] // for temporary image storage
4. For (i=0; i<=imgArr; i++)
5. do
6. separate gray and white pixels and store in grArr[] and whArr[]
7. if (invalid pixel)
8. then segmentation result is invalid
9. else
10. feed to CNN model layer
11. end

CNN's are found in image distinction, graphic acknowledgement, and object monitoring complications, as well as , carry large activities. Conventional image classification techniques generally include stages such as for example attribute diagnosis, characteristic removal, feature decrease, and so classification. Involving all these stages, characteristic recognition and extraction will be the most crucial phases when it comes to distinction effectiveness. Therefore, we recommended deep neural network to execute brain Large as well as, Low Quality Glioma (HGG/LGG) tumor segmentation. With a convolution by the end of TumorHunt algorithm we've an productive execution, the prediction at check time for a complete brain is captured 10 times quicker.

III. RESULTS AND DISCUSSION

In this study, we presented proposed research, a CNN-based delineation method designed for automatic brain tumor segmentation for early detection. To compare with other algorithms the developed Auto-Aliasing strategy and trained on multi-modality BRATS data. The competitive results obtained from BRATS2013 are shown in Table. The developed method ranked first in the tumor region factors analysis. This result indicates that proposed method performs well on multimodality image segmentation.

Parameters Tested	Proposed System Results	Existing Results
DSC	90.734	88.182
PPV	0.789	0.540
Accuracy	89.870	82.562

IV. CONCLUSION

As there is a need of automatic brain tumor segmentation, this paper described the new machine learning algorithm with deep CNN model training and testing. The proposed research is efficient to process MR images in a very fast way. This will provide quick diagnosis solution for patients. As machine learning algorithms are need more time for preprocessing, we used Gabor filter for processing of initial input set of images. This system further can be developed for more precise outcome using Artificial Intelligence support.

V. REFERENCES

[1]. Jaglan, Poonam, Rajeshwar Dass, and Manoj Duhan. "A comparative analysis of various image segmentation techniques." Proceedings of 2nd International Conference on Communication, Computing and Networking. Springer, Singapore, 2019.

[2]. Shirly, S., and K. Ramesh. "Review on 2D and 3D MRI image segmentation techniques." Current Medical Imaging 15.2 (2019): 150 160.

[3]. Borne, Léonie, et al. "Automatic labeling of cortical sulci using patch-or CNN-based segmentation techniques combined with bottom-up geometric constraints." Medical Image Analysis (2020): 101651.

[4]. Kassim, Siti Rafidah Binti, et al. "Implementation of Image Segmentation Techniques to Detect MRI Glioma Tumour: Mid-Level Image-Processing." 2019 6th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI). IEEE, 2019.

[5]. Zhuang, Zhemin, et al. "Application of fractal theory and fuzzy enhancement in ultrasound image segmentation." Medical & biological engineering & computing 57.3 (2019): 623-632.

[6]. Stewart, Jebb, et al. "Understanding What the Machine Sees When Using Deep Learning for Image Segmentation to Detect Convection Initiation." AGUFM 2019 (2019): A53H-07.

[7]. van Sloun, Ruud JG, et al. "Deep learning for real-time, automatic, and scanner-adapted prostate (zone) segmentation of transrectal ultrasound, for example, magnetic resonance imaging–transrectal ultrasound fusion prostate biopsy." European urology focus (2019).

[8]. Han, Seungwook, et al. "3D distributed deep learning framework for prediction of human intelligence from brain MRI." Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging. Vol. 11317. International Society for Optics and Imagenics, 2020.

[9]. Sharma, Pallabi, et al. "Classification of Brain MRI Using Deep Learning Techniques." Soft Computing: Theories and Applications. Springer, Singapore, 2020. 559-569.

[10]. Sujit, Sheeba J., et al. "Automated image quality evaluation of structural brain MRI using an ensemble of deep learning networks." Journal of Magnetic Resonance Imaging 50.4 (2019): 1260-1267.

[11]. Xiao, Zhe, et al. "A deep learning-based segmentation method for brain tumor in MR images." 2016 IEEE 6th International Conference on Computational Advances in Bio and Medical Sciences (ICCABS). IEEE, 2016.

[12]. Amin, Javaria, et al. "Brain tumor classification based on DWT fusion of MRI sequences using convolutional neural network." Pattern Recognition Letters 129 (2020): 115-122.

[13]. Hollon, Todd C., et al. "Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks." Nature Medicine (2020): 1-7.

[14]. Kang, Ruirui, et al. "Prior information constrained alternating direction method of multipliers for longitudinal compressive sensing MR imaging." Neurocomputing 376 (2020): 128-140.

[15]. Chaudhary, Atish, and Vandana Bhattacharjee. "An efficient method for brain tumor detection and categorization using MRI images by K-means clustering & DWT." International Journal of Information Technology 12.1 (2020): 141-148.

[16]. Gandhi, Meet, Juhi Kamdar, and Manan Shah. "Preprocessing of Non-symmetrical Images for Edge Detection." Augmented Human Research 5.1 (2020): 10.

[17]. Xiao, Zhikang, Yang Zou, and Zhen Wang. "An improved dynamic double threshold Canny edge detection algorithm." MIPPR 2019: Pattern Recognition and Computer Vision. Vol. 11430. International Society for Optics and Imagenics, 2020.

[18]. Khan, Hikmat, et al. "Cascading handcrafted features and Convolutional Neural Network for IoT-enabled brain tumor segmentation." Computer Communications (2020).

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

A Survey on Healthcare Assistant using Machine Learning and NLP Models

Shivani Bahirat[1], Saniya Bagwan[1], Nupur Dhamal[1], Sagar Patil[1] , Prof. Bhagyashree Dhakulkar[2]

[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

[2]Professor, Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

The Healthcare conveyance framework is cost-restrictive, inefficient, and unsustainable in the current situation. Supporting this, Machine Learning (ML) has changed the manner in which associations and people use the information to improve the proficiency of a framework.AI calculations permit planners to manage an assortment of organized, unstructured, and semi-organized information. In this survey, various applications of Machine Learning and Natural Language Processing to ease the functionality of the health-care system was included. With the increase in workforce shortages and discontent in healthcare services, there is a need to provide assistant to professionals and normal people to ease the facilities and services on time, cost-effective, and efficiently at the local level. Artificial Intelligence is used to predict diseases based on a list of symptoms and provide treatment available. This will minimize the gap between service providers and localists.

Keywords: Natural Language Processing, Machine Learning, Health Assistant, Artificial Intelligence.

I. INTRODUCTION

Advances in current innovations have caused a move towards computerized well-being in medical services, where clinical and authoritative exercises can be helped by Machine Learning algorithms and Natural Language Processing Models to provide healthcare services at ease. This requires efficient utilization of accurate medical records for analysis. An automatized Healthcare assistant chatbot system is a framework with human communication utilizing normal language analysis to give the basic medical guide to save time and money.

Technical devices need a stage where they can comprehend the essential language of people and decipher it in a reasonable structure. Assuming a sound clip of an individual or individuals talking, determine the textual representation of the speech. This is something, contrary to text to discourse and is one of the amazingly troublesome issues informally named AI-complete. In normal discourse, there are not really any stops between progressive words, and thus speech segmentation is a necessary subtask of speech recognition. These models are used to interact with users to provide healthcare and medical guidelines.

It has been discovered that the expense of getting a regular checkup from medical professionals, for an hour is expanded. Individuals don't discover time to look after their health. For clients with such a

bustling timetable, the healthcare bot gives major help. Medical services towards this path will convey incredible advantages to society, acquiring material upgrades normal life expectancies, and personal satisfaction. This will welcome on advanced development of the wellbeing area, described by the capacity to rapidly diagnose, identify emergency clinics and the professionals that are best prepared to treat the condition. Machine learning (ML), an important application of artificial intelligence (AI) that focuses on the advancement of computer programs that can access data and use it to train for themselves. The use of machine learning algorithms will produce prediction and recommendation based desired output for a particular system. The objective of Machine Learning in healthcare is very perplexing and requesting. Finding exact therapy alternatives for an individual dependent on their own medical history, lifestyle routines, hereditary information, and persistently changing neurotic tests. Normally, we have to bring the most remarkable AI procedures — profound Machine Learning algorithms, AI-driven inquiry calculations/recommendation's fortification learning, probabilistic graphical models, semi-directed learning to handle the issues.

II. LITERATURE REVIEW

The zone of Computer Science that centers around creating advances to improve wellbeing, prosperity, and medical care is generally known as eHealth. E-Health can be defined as a way to interact with patients using technology. The aim is to assist the patients by lessening the complexity of their most basic tasks in the health domain, so as to add to the public medical services framework in a roundabout way. The Systems surveyed are described underneath.

Divya Madhu et al. (2017) [1] proposed a system for predicting the disease based on symptom analysis and provides output in form of available treatment. The approach included Medical Assistance using a trained chatbot.

The chatbot is built on different terms such as machine translation, virtual agent, dialogue system, and Chatterbot. The system also contained Medical details so as to deliver the desired information to the user whenever required. The proposed model is a simple and real-time chat system that works in cross-platform devices and is easily integrated and upgradable. The medicine dosage recommended is based on age and weight provided by the user.

Amruta Kulkarni et al. (2017) [2] elaborates on the prediction of diabetes and displays the use of an android device as a front end to store the collected data in cloud. The data stored in the cloud is used for accurate prediction of diabetes. The system relies on stored health parameters collected using heart sensors, glucose meter, etc. The service of the system is accessed using utilities of an android device such as a speaker, microphone, etc. The major algorithms used in prediction were Random Forest Algorithm and Logistic Regression. The predictive model analyses diabetes using stored data and predicts output with an accuracy of 80.01%. The design of the system satisfactorily eases its use for especially older patients.

Sushree Satapathy et al. (2018) [4] proposed a Geographic Information System (GIS) based system that aims at providing information about nearest healthcare centers, helpline number in emergencies, prior first-aid knowledge, and availability of transport for a medical emergency by using technologies like machine learning and Geographic Information System (GIS). The main focus was to reduce the response time during medical emergencies and save lives by providing a platform for collaboration with caregivers like Clinics, Blood bank, Ambulance Services, etc. The prominent technologies used in the

system are Mapbox GL JS to manipulate spatial data and NodeJS to develop the back end along with the PostgreSQL database. The proposed system comprises three modules for three different stakeholders namely, User, Hospital, and Ambulance. Due to the presence of various stakeholders, all the essential medical emergency data is provided in a single application for easy usage.

Muhammad Azeem Sarwar et al. (2018) [3] discusses the predictive analytics in healthcare by using six different machine learning algorithms on a specified dataset. The research also include comparison of algorithms in terms of accuracy and performance. The main goal of analysis is to ease the prediction of diabetes for doctors and practitioners using Machine learning techniques. Among various algorithms like SVM, KNN, LR, DT, RF, and NB, SVM and KNN gives highest accuracy for predicting the Indian dataset consisting 768 records. The both algorithms provide 77% accuracy in comparison with other techniques.

NaliniPriya G et al. (2019) [5] presents a smart personal health care assistant that tracks one's activities, moods and suggests precautious actions accordingly. The system makes use of personal assistance chatbot to manage and revert the messages from the users by using machine learning techniques. The user can track their own lifestyle through the use of a system which will lead to a healthier life. The system comprises a wearable device containing sensors along with the mobile phone, connected via Bluetooth. This helps to keep a measure of body temperature, heart-beat rate, body movement, and breathe rate. The sensors embedded in the wearable devices sense the temperature and other parameters of a person and send the data to the control center. The system has the ability to capture the user's voice message to revert it by matching the input with the database used. The system deals with Analog as well as Discrete data. The system is based on Logistic regression to provide an accuracy of 0.96 more than other machine learning algorithms.

Moshuir Rahman et al. (2019) [6] demonstrates the implementation of Healthcare chatbot in the Bangla language. The system makes use of Machine Learning algorithms and customized Bangla datasets. The demonstration includes the use of six various supervised Machine Learning algorithms, concluding SVM for best performance. The system collects the user's basic details in the preferred language in textual form, where the Name Entity Recognition algorithm is used to extract the name from the input provided by the user.
The later phase consists of two basic commands mainly, Disease classification command and General command. The Disease classification command collects symptoms from the user and generates appropriate results based on training and testing dataset. Whereas in General command the user is provided with the information related to questions asked regarding health issues and its classification. The accuracy of the system is maintained by the use of the SVM algorithm.

Prakhar Srivastava et al. (2020) [7] uses AIML (Artificial Intelligence Mark-up Language) to withdraw the patterns of messages. The symptoms extracted from the pattern snippets of the messages from the user's conversation are matched with the database and further sent to the bot engine for prediction and diagnosis. Algorithms like K-nearest neighbor (KNN) and SVM hold good command over complex classification tasks to split the data in training and testing sets, so as to accurately classify the disease and pass on the suggestions. The bot makes use of a closed-source, external recognition engine to revive the functionality of diagnosis.

Zhibin Liao, Qi Wu et al. (2020) [8] proposed a Medical Data Enquiry system using a Question Answering model based on Natural Language Processing. The QA system was developed in PyTorch. The system built is for health professionals which are based on SQLOVA models and BERT. The system was fine-tuned using a dataset for 100 epochs, also the Bilinear similarity model is trained with the Adam optimizer by reducing the learning rate after epochs. The system was trained initially by 75 real questions and 455 back-translated questions with a retrieval accuracy of 78%.

Marco Polignano, Fedelucio Narducci et al. (2020) [9] presents a system containing HealthAssistantBot (HAB) along with different modules. Initially, the user can create a profile, and the system identifies the condition of the user by System Checker (SC). These symptoms are analyzed and passed on to Recommender System (RS) where Machine learning algorithms are used to provide the outcome. The Virtual assistant processes the user's input using Natural Language Processing models. To implement this intuitive process, the team designed the assistant as a Conversational Agent (CA) to interact with users in natural language. The CA comprises of Intent Recognition (IR) and Entity recognition (EI). The classification of disease in the Recommender system module is implemented using manual clustering of the dataset. The system gained a success ratio of 76.271% on real use cases.

Lekha Athota et al. (2020) [9] proposed the system based on Artificial Intelligence that diagnose the disease and display basic details about the disease before taking professional medical consult. The system is basically a chatbot that stores the database to recognise the question and makes query decision to answer the question. Ranking and question similiarity calculation is is estimated using n-gram, TF-IDF and cosine similarity.

III.CONCLUSION

This paper represents an analysis of the existing use of Machine Learning algorithms and Natural Language Processing models to decrease the time lapse between the health service and end-user to perfect the medical profession. Customized chatbot assistants, to provide medical care represents few difficulties, as the precision of result and implementation of efficient algorithms could ease the gap between doctors and patients today. By analyzing the above algorithms, the use of efficient algorithms like KNN, SVM, and Decision Tree achieves precision. This approach understands the need of a patient (user) before the need of the organization, to provide health care efficiently and on time. The accuracy is increased by making use of an appropriate dataset, and training the dataset to yield the precise result and prediction. According to the survey, efficient algorithms and accurate datasets lead the system to a safe and successful health-care assistance system.

IV. REFERENCES

[1]. Divya Madhu, Neeraj Jain C. J, Elmy Sebastain, Shinoy Shaji, and Anandhu Ajayakumar. (2017) "A Novel Approach for Medical Assistance using Trained Chatbot", International Conference of Inventive Communication and Computational Technologies (ICICCT).

[2]. Amruta Kulkarni, Deepa Kalburgi, and Poonam Ghuli (2017). "Design of Predictive Model for Healthcare Assistance Using Voice Recognition", 2nd IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions 2017.

[3]. Muhammad Azeem Sarwar, Nasir Kamal, Wajeeha Hamid, and Munam Ali Shah (2018). "Prediction of Diabetes Using Machine Learning Algorithm in Healthcare", Proceedings of the 24th International Conference on Automation and Computing NewCastle University, UK, 6-7 September 2018.

[4]. Sushree Satapathy, Atashi Khatua, Jennifer Rodriques, Madhulika Tadas, and Premlata Pawar (2019). "Sahaay – A medical assistance system", Global Conference for Advancement Technology (GCAT), Bangalore, India. Oct 18-20, 2019.

[5]. NaliniPriya G, Priyadarshani P, Puja Shree S, and RajaRajeshwari K (2019). "BayMax: A Smart Halthcare System provide services to Millennials using Machine learning technique", IEEE 6th International Conference on smart structures and systems ICSSS 2019.

[6]. Moshuir Rahman, Ruhul Amin, Nazmul Khan Liton, and Nahid Hossain (2019). "Disha: An implementation of Machine Learning based Bangla Healthcare Chatbot", 22nd International Conference of Computer and Information Technology (ICCIT), 18-20 December, 2019.

[7]. Prakhar Srivastava, and Nishant Singh (2020). "Automized Medical Chatbot (Medi-bot)", International Conference on Power Electronic & IoT Applications in Renewable Energy and its Control (PARC) GLA University, Matura, UP, India. Feb 28-29, 2020.

[8]. Zhibin Liao, Lingqiao Liu, Qi Wu, Damien Teney, and Chunhua Shen (2020). "Medical Data Inquiry using a Question Answering Model", IEEE 17th International Symposium on Biomedical Imaging (ISBI), lowa, USA 2020.

[9]. Marco Plignano, Fedelucio Narducci, Andrea Iovine, Cataldo Musto, Marco De Gemmis, and Giovanni Seneraro (2020). "HealthAssistantBot: A personal health assistant for the Italian Language", Digital Object Identifier IEEE Access, May 26, 2020.

[10]. Lekha Athota, Vinod Kumar Shukla, Nitin Pandey, and Ajay Rana (2020). "Chatbot for Healthcare System Using Artificial Intelligence", 8th International Conference on Reliability, Infocom Technologies and Optimization(Trends and Future Directions) (ICRITO) Amity University, Noida, India. June 4-5, 2020.

[11]. Tejasvi Wani, and Bhagyashree Dhakulkar (2018). " Survey on auditable health records levering DROPS in cloud, Open Access International Journal of Science and Engineering (OAIJSE), volume 3, issue 9, September 2018.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Career Path Prediction Using Machine Learning

Prathamesh Gavhane[1], Dhanraj Shinde[1], Ashwini Lomte[1], Naveen Nattuva[1], Monika Munjal[1]
[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

In today's era, choosing the right career option is the challenging task. Starting at the early stage of life students usually fail to grasp the idea of which career to pursue as they lack maturity and the experience related to that field. Furthermore, students suffer greatly in deciding which career would result the highest benefit. Students do not have sufficient knowledge to take the decision on their own which may lead to complications in future. Choosing proper career can help the student to be successful by excelling his/her skills in that particular field. In this paper we present the detail study for choosing the right career option for an individual. Survey has been done for analyzing the different techniques for making right career choice.

Keywords: Machine Learning, Classification algorithms, Career path prediction.

I. INTRODUCTION

Machine Learning (ML) is the study of computer algorithms that improve automatically through experience. It is seen as a subset of artificial intelligence. Machine learning algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications. Machine learning helps the computers to act without explicitly being programmed. Simply it is giving computers the ability to learn by using statistical techniques [11]. This helps in solving very complex tasks and problems very easily and without involving much human labor.

Nowadays students are getting confused about their right career path. Because of this confusion they are ending up with the career in which they are least interested. In today's world competition is heavily increasing day by day. Mainly it is too heavy in present day's technical society. Students need to be well organized and planned from initial stages of their education, so as to reach the goal. To help them in improving themselves, motivating themselves to a better career path. It is very important to constantly evaluate their performance, identify their interests and evaluate how close they are to their goal and whether they are in the right path that directs towards their targeted [11].

There are many new career opportunities in every field, because with the increase in research and exploration in various domains. The reasons for this confusion could be unawareness of self-talent and self-personality trait, unawareness of the various options available, equal interests in multiple fields, less exposure, market boom, assumed social life, peer-pressure etc. This creates more confusion to the students to select one career option. There should be proper counseling of the student's psychology,

interest and their capacity to work in a particular field [7]. Otherwise, student may select a wrong career option and the consequences of this wrong decision could be work dissatisfaction, poor performance, anxiety and stress, social disregard etc.

There are career counseling services which are helping students to find their career goals, which is the reason counseling centers having been established. These counseling centers helps student to know the wide variety of options available for them. Now students can choose the best path for them provided by the counsellor.

In this paper we are going to provide a machine learning model which will give you the career path prediction. To the best of our knowledge, there is no available benchmark dataset suitable for career path modeling [3]. We thus created new datasets by crawling fifteen popular career paths, namely engineer, doctor, pharmacist, lawyer, archaeologist, financial advisor, motivational speaker, chartered accountant, hotel management, wedding planner, writer, photographer, cabin crew, journalist and translator. For each career path we have an individual dataset. We have collected the dataset by forwarding the google form links to the students, teachers, professional which were containing some questions to answer. Likewise, we successfully completed the collection of datasets.

II. LITERATURE REVIEW

[1] Roshani Ade & P. R. Deshmukh (2014). In this paper for classification of students using psychometric tests. They used incremental naive bayes algorithm. And the results were TP-Rate_0.896, FP Rate_0.01, Precision_0.903, Recall_0.896, F-Measure_0.893 and ROC-Area_0.99. In future naïve bayes algorithm can be used as a weak classifier in the ensemble concept for incremental learning.

[2] Ahmad F. Subahi (2018). He proposes a data collection strategy to build the required career path prediction dataset for a promising data-driven system. A new artificial neural network (ANN) approach for career path prediction was used.

[3] Ye Liu, ET AL (2016). They have created a career path prediction model for career path instead of going to the fortune tellers. They have collected the information from various social networks. And the future work is to extend the model to consider the source descriptiveness and learn the source confidence adaptively.

[4] Beth Dietz-Uhler & Janet E. Hurn (2013). So, they have used a learning analytics to predict student success through a perspective of faculty. In this paper, they defined about learning analytics, how educational institutions has been used it, what learning analytics tools are available and how faculty can make use of data in their courser to improve the performance of students.

[5] Min Nie, ET AL (2020). In past, professional career appraisers used questionnaires to suggest the best career path for a student, instead of that they have created a career choice prediction based on campus big data mining the potential behavior of college students. Algorithm used is XGBOOST (ACCBOX). Accuracy of ACCBOX was 0.638.

[6] Amer Al-Badarenah & Jamal Alsakran (2016). As we know that there are recommendation systems for the recommendation purpose while online shopping, movies, songs, etc. In that way they have created an automated recommender system for course selection which will be easy for students to choose the right subject for them.

[7] Nikita Gorad, ET AL (2017). Keeping in mind that selecting the right career is one of the important decisions. Some students end up

selecting wrong decision. For that purpose, they have created a career counselling model using data mining. They used adaptive boosting algorithm which gave around 94% of accuracy.

[8] Dileep Chaudhary, ET AL (2019). For selecting an appropriate career path, they have created a student future prediction model using machine learning. Algorithms used were linear regression, decision tree and random forest, to improve accuracy they used adaptive boosting over the algorithms.

[9] Vivek Kumar Mourya, ET AL (2020). They have created a career guide application using machine learning. Through this application students can easily choose a best career path for them. The machine learning algorithm used for predicting is a clustering algorithm named as K-means algorithm.

[10] Lakshmi Prasanna & DR.D.Haritha (2019). Keeping recommender system in mind, they have created a smart career guidance and recommendation system. This paper proposes feasible predictions for student's field selection based on their marks and choice of interest. Ten to eleven machine learning algorithms were used for the predictions. In which logistic regression gave 82% accuracy. In future we can use clustering methods for better understanding.

[11] K. Sripath Roy, ET AL (2018). They have created a student career prediction model using advanced machine learning techniques. Algorithms used are support vector machine (SVM), xgboost and decision tree. SVM gave more accuracy with 90.3 percent and then the XG Boost with 88.33 percent accuracy.

[12] Mubarak Albarka Umar (2019). A case study of student academic performance prediction using artificial neural networks was presented. This study presents a neural network model capable of predicting student's GPA using students' personal information, academic information, and place of residence. Thus, the model correctly predicts 73.68% of student performance and specifically, 66.67% of students that are likely to dropout or experience delay before graduating.

[13] Ezenkw.C.P, ET AL (2017). In this paper, an Automated Career Guidance Expert System (AC-GES) has been developed using case-based reasoning (CBR) technique. AC-GES is to assist high school students in choosing career paths that best suit their abilities based on their previous performances in some selected subjects, using Nigerian students as a case study.

[14] Elaf Abu Amrieh, ET AL (2016). They have used data mining technique in educational data to predict student's academic performance using ensemble methods. They have used bagging, boosting and random forest (RF) and set of classifiers such as artificial neural network, naïve bayesian and decision tree. The obtained results reveal that there is a strong relationship between learner's behaviors and their academic achievement.

[15] Sudheep Elayidom, ET AL (2009). They have applied data mining on dataset using statistical techniques for career selection. This will help the students in a great way in deciding the right path for them for a bright future. The software developed is simple to use besides being reasonably accurate. Moreover, the user-friendly interface used in this project turns out to be easy to handle and avoid complications.

[16] Maha Nawaz, ET AL (2014). In this paper they have created an automated career counseling system for students using case-based reasoning (CBR) and J48. This model presents an automated system that copies a one-to-one meeting with a professional career counselor. Out of the two algorithms tested, CBR gave the highest accuracy and Decision tree J-48 gave the

lowest accuracy. The results indicate that the system is capable of correctly proposing majors with approximately 80% accuracy when presented with sufficient data and features.

III. CONCLUSION

Career path prediction is essential as it helps to identify the interest, self-talent, and potential of the student. If the student takes the wrong decision in the early stage of his/her career he/she will face a lot many problems like lack of interest, leaving the course uncompleted, frustration, wastage of time and money etc. It is very important to build reliable models which can predict career path so that the student can escape huge loss and it is also difficult to start with the new career option in this competitive world. The paper presents a review of career path prediction. It projects many attributes and techniques used to predict career path. The purpose of the paper is not to introduce a new technique but to review the implementation and understanding of the existing models. In this paper, 16 techniques have been discussed. From the survey conducted it is understood that predicting career path is important for student.

IV. REFERENCES

[1]. Ade R. and Deshmukh P. R. (2014). Classification of Students Using psychometric tests with the help of Incremental Naive Bayes Algorithm. International Journal of Computer Applications. (0975 – 8887) Volume 89 – No 14.

[2]. Subahi A., F. (2018). Data Collection for Career Path Prediction Based on Analyzing Body of Knowledge of Computer Science Degrees. Journal of Software. Volume 13.

[3]. Liu Y., Zhang L., Nie L., Yan Y., Rosenblum D. S. (2016). Fortune Teller: Predicting Your Career Path. Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16).

[4]. Uhler B. D., Hurn J. E. (2013). Using Learning Analytics to Predict (and Improve) Student Success: A Faculty Perspective. Journal of Interactive Online Learning. Volume 12.

[5]. Nie M., Xiong Z., Zhong R., Deng W., Yang G. (2020). Career Choice Prediction Based on Campus Big Data—Mining the Potential Behavior of College Students. Applied science. a.Doi: 10.3390/app10082841.

[6]. Badarenah A. A., Alsakran J. (2016). An Automated Recommender System for Course Selection. International Journal of Advanced Computer Science and Applications, Vol. 7, No. 3.

[7]. Gorad N., Zalte I., Nandi A., Nayak D. (2017). Career Counselling Using Data Mining. International Journal of Innovative Research in Computer and Communication Engineering. Vol. 5, Issue 4.

[8]. Chaudhary D., Prajapati H., Rathod R., Patel P., Gurjwar R. K. (2019). Student Future Prediction Using Machine Learning. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. Volume 5, Issue 2.

[9]. Mourya V., Phatale S., Thakur S., Mane P. (2020). Career Guide Application using ML. International Research Journal of Engineering and Technology (IRJET). Volume: 07 Issue: 09.

[10]. Prasanna L., Haritha D. (2019). Smart Career Guidance and Recommendation System. International Journal of Engineering Development and Research. Volume 7, Issue 3.

[11]. Roy K. S., Roopkanth K., Uday V., Bhavana V., Priyanka J. (2018). Student Career Prediction Using Advanced Machine Learning Techniques. International Journal of Engineering & Technology.

[12]. Umar M. A. (2019). Student Academic Performance Prediction using Artificial Neural Networks: A Case Study. International Journal of Computer Applications (0975 – 8887) Volume 178.

[13]. Ezenkwu C.P., Johnson E.H., Jerome O.B. (2017). Automated Career Guidance Expert System Using Case-Based Reasoning Technique. Cisd ijournal. Volume 8, No. 1.

[14]. Amieh E. A., Hamtini T., Aljarah I. (2016). Mining Educational Data to Predict Student's academic Performance using Ensemble Methods. International Journal of Database Theory and Application. doi.org/10.14257/ijdta.2016.9.8.13.

[15]. Elayidom S., Idikkula S. M., Alexander J. (2009). Applying Data mining using Statistical Techniques for Career Selection. International Journal of Recent Trends in Engineering, Vol. 1, No. 1.

[16]. Nawaz M., Adnan A., Tariq U., Salman J. F., Asjad R., Tamoor M. (2014). Automated Career Counseling System for Students using CBR and J48. Journal of Applied Environmental and Biological Sciences.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

A Survey on Crop yield prediction using Machine Learning

Shivam Parande[1], Muskan Mulani[1], Meghashree Choudhary[1], Neha Singh[1], Prof. Aseema Jana[2]

[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

[2]Professor, Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

Machine Learning is an emerging research field in crop yield analysis and crop prediction analysis too. Yield prediction and crop prediction are very important issue in agricultural. India being a farming nation, its economy transcendently relies upon farming yield development and agroindustry items. Ranchers for the most part have the outlook of planting a similar yield, utilizing more composts and following the public decision. The normal issues the vast majority of the ranchers confronting today is about forecast of harvest and yield. Any farmer will be interested to know which crop to grow, so that it increases its yield production. This paper describes various methods and algorithm's used for crop prediction and crop yield prediction.

Keywords: Climate, Machine Learning, Algorithms, Prediction.

I. INTRODUCTION

The primary objective of agricultural arranging and planning is to accomplish most extreme yield pace of harvests by utilizing set number of land assets. Many AI calculations can help in improving the creation of harvest yield rate. At whatever point there is misfortune in negative conditions. We can apply crop choosing technique and diminish the misfortunes. Also, it tends to be utilized to pick up crop yield rate in ideal conditions. This boosting of yield rate helps in improving nation economy. We have a portion of the elements that impact the crop yield rate. They are seed quality and yield choice. We need to test the nature of the seeds prior to planting. As we realize that great nature of seeds helps in getting more yield rate. Also, determination of harvests relies on two things that is positive and troublesome conditions. This can likewise be improved by utilizing hybridization techniques. Numerous explores are completed to improve horticultural arranging. The objective is to get the greatest yield of harvests. Numerous order strategies are likewise applied to get most extreme yield of harvests. AI methods can be utilized to improve the yield pace of harvests. The technique for crop determination is applied improve crop creation. The creation of yields may rely upon geological states of the district like stream ground, slope zones or the profundity zones. Climate conditions like dampness, precipitation, temperature, cloud. Soil type might be dirt, sandy, saline or peaty. Soil organization can be copper, potassium, phosphate, nitrogen, manganese, iron, calcium, ph. esteem or carbon and various strategies for gathering. Numerous

boundaries are utilized for various yields to do extraordinary expectations. These expectation models can be concentrated by utilizing explores. These forecasts are named two types. One is traditional measurement technique and other is ML strategies. Traditional technique helps in foreseeing single example spaces. Furthermore, ML techniques helps in anticipating various forecasts. We need not to consider the structure of information models in traditional strategy where as we have to think about the structure of information models in Machine Learning techniques.

AI (ML) is the investigation of computer calculations that improve naturally through experience. It is viewed as a subset of man-made reasoning. AI calculations assemble a model dependent on test information, known as "preparing information", to settle on forecasts or choices without being unequivocally customized to do as such. AI calculations are utilized in a wide assortment of uses, for example, email sifting and PC vision, where it is troublesome or impractical to create traditional calculations to play out the required assignments. A subset of AI is firmly identified with computational insights, which centres around making predictions utilizing computer systems; yet not all AI is measurable learning. The investigation of numerical streamlining conveys strategies, hypothesis and application spaces to the field of AI. Information mining is a connected field of study, zeroing in on exploratory information investigation through unaided learning. In its application across business issues, AI is likewise alluded to as predictive investigation.

II. LITERATURE SURVEY

- **Shrikant Kokate** et al. 2020 [1] proposed a crop recommendation system using classifier models. The proposed framework is versatile as it very well may be utilized to test on various harvests. This project was basically made by taking a particular city into consideration and predicting the suitable crop for the area which falls under that city. The data set used had different modules such as rainfall prediction, weather report and crop section. Soil parameter contained soil type and soil ph value. The weather report was comprised of humidity, temperature, wind and rainfall. Dataset also had the information of crop production in the last 10 years in that particular area. The algorithms and techniques used were k-nearest neighbour and decision tree algorithm. A comparative analysis was made between the two and it was found that accuracy of decision tree algorithm was 76.8% and that of KNN algorithm was 89.4%. Naïve Bayes classifier model was also used which had an average accuracy of 70-75%. The results showed that combination classification algorithm like naïve bayes and decision tree classifier are better performing than use of single classifier model.
- **S.Bhanumati** et al. 2019 [2] proposed a model which predicted the crop yield production and efficient use of fertilizers. The proposed system predicted the result which will be helpful for farmers to know its overall yield of various crops so that he can choose the crop which will be giving highest yield. The system also tells the efficient use of fertilizer so that he can use only the required amount of fertilizer for the required field. The dataset contained parameters such as state name, district name, crop, area, season, etc. A comparative analysis was done between random forest algorithm and back propagation model. Random forest algorithm showed good results. Even the fertilizer data was trained using back propagation algorithm. As suggested a web application can be made to deploy the model in the future and make it user friendly.

- **Suvidha Jambekar** et al 2018 [3] have proposed a system of prediction of crop prop production in India using data mining techniques. The dataset used was collected from the public ally available records of the Indian government for the duration of 64 years from 1950-2013. It consists of monthly rainfall, monthly mean temperature, area under irrigation, production. Regression analysis was used for prediction of crop production. Namely three regression algorithms like random forest, multiple regression and multivariate adaptive regression spline were used. The concluded results showed that the performance of multivariate adaptive regression splines was better as compared to multiple linear regression and random forest algorithm.
- **Niketa** et al 2016 [4] have shown that the yield of the crop relies upon the occasional atmosphere. In India, atmosphere conditions shift unequivocally. In the hour of dry season, ranchers face major issues. So this mulled over they utilized some machine learning calculations to help the ranchers to recommend the harvest for the better yield. They take different information from the earlier years to assess future information. They utilized some SMO classifiers in WEKA to arrange the outcomes. The fundamental calculates that take thought are least temperature, greatest temperature, normal temperature, and earlier year's harvest data and yield data. Utilizing SMO device they arranged the past information into two classes that are high return and low yield. The acquired outcome for the crop yield expectation utilizing SMO classifier gives less exactness when contrasted with guileless Bayes, multilayer perceptron and Bayesian organization.
- **Eswari** et al 2018 [5] have shown that yield of the harvest relies upon the discernment, normal, least and greatest temperature. Separated, from that, they have taken one more quality named crop evapotranspiration. The yield evapotranspiration is an element of both the climate and development phase of the plant. This trait is taken into thought to get a decent choice on the yield of the gatherings. They all gathered the dataset with these qualities and send as contribution to the Bayesian organize and group into the two classes named valid and bogus classes and contrasted and the watched orders in the model with a disarray network and bring the precision. At last, they presumed that harvest yield expectation with Naïve Bayes and Bayesian organization give high precision when contrasted with SMO classifier and gauging the harvest yield expectation in various atmosphere and trimming situations will be valuable.
- **Parul Agarwal** et al 2019 [6] proposed a model for prediction crop yield prediction using machine learning algorithms. The main aim of the built system was to help choose the right crop to the farmer. The dataset used was collected online from official website of Indian government. The dataset comprised of temperature, rainfall and production of previous years. Various machine learning algorithms such as random forest algorithm, XG Boost classifier, logistic regression and KNN classifier were used. The results showed that Random forest algorithm had the highest accuracy.
- **J.P. Singh** et al 2015 [7] have proposed a system in their paper which helps in improving the yield pace of harvests by applying grouping strategies and contrasting the boundaries. We can likewise do breaking down and expectation of harvests utilizing baysian calculations. The calculations utilized are Bayesian calculation, K-implies Calculation, Clustering Algorithm, Support Vector Machine.
- **Aditi Kharde** et al 2018 [8] have inferred that the paper helps in foreseeing crop arrangements and

augmenting yield rates and making advantages to the ranchers. Additionally, Using Machine Learning applications with farming in anticipating crop sicknesses, examining crop re-enactments, distinctive water system designs. IOT devices are used for temperature sensing, humidity sensing. An android application is built for displaying the required information. The calculations utilized are K-means Clustering.

III.CONCLUSION

This paper represents analysis of existing use of Machine Learning algorithms and Deep Learning algorithms for helping the farmer in choosing an efficient crop for his land. This paper describes various methods and algorithm's used for crop prediction and crop yield prediction. By analysing the above algorithms, use of efficient algorithms like KNN Algorithm, SVM, and Decision Tree achieves the precision. According to survey, efficient algorithms and accurate dataset lead to proper prediction of yield and crop.

IV. REFERENCES

[1]. Bhanumathi, M.Vineeth, N.Rohit (2019). "Crop Yield Prediction and Efficient use of Fertilizers", IEEE International Conference on Communication and Signal Processing, April 4-6 2019, India.

[2]. Suvidha Jambekar, Shikha Nema, Zia Saquib (2018). "Prediction of Crop Production In India Using Data Mining Techniques", IEEE Fourth International conference on Computing Communication Control And Automation, 2018.

[3]. Niketa Gandhi, Leisa Armstrong, Owaiz Petkar, Amiya Kumar Tripathy (2016). "Rice crop yield prediction in India using support vector machines" IEEE Conference, 2016.

[4]. K. E. Eswari, L.Vinitha (2018). "Crop Yield Prediction in Tamil Nadu using Bayesian Network", International Journal of Intellectual Advancements and Research in Engineering Computations, 2016.

[5]. Pavan Patil, Virendra Panpatil, Prof.Shrikant Kokate (2020). "Crop Prediction Systems Using Machine Learning", International Research Journal Of Engineering and Technology, 02 Feb 2020.

[6]. Aruvansh Nigam, Saksham Garg, Archit Agrawal, Parul Agrawal (2019). "Crop Yield Prediction Using Machine Learning Algorithms", Fifth IEEE International Conference on Image Information Processing, 2019.

[7]. Rakesh Kumar; M.P. Singh; Prabhat Kumar; J.P. Singh (2015). "Crop Selection Method to maximize crop yield rate using machine learning technique", IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), 2015.

[8]. K.D.Yesugade, Aditi Kharde , Ketki Mirashi , Kajal Muley ,Hetanshi Chudasama (2018). "Machine Learning Approach for Crop Selection based on Agro-Climatic Conditions", International Journal of Advanced Research in Computer and Communication Engineering, 2018.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Product Grading System Using Blockchain

Yash Kothari[1], Debajit Roy[1], Sanskriti Gupta[1], Vishwajeet George[1], Aseema Jana[1]
[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

Ratings and Reviews affect our decision making drastically, the current review systems are usually centralized and can be tampered. In today's era, choosing the right product has become an arduous task. So many options are available in the market, so many products are being advertised which make similar claims that the buyer gets confused which one is true and should be relied upon. It breeds a lot of dilemma in the mind of the customer. Essentially, a review helps to get the insight of the product, typically it describes what sets the product apart from others. A review offers valuable guidance from people who have used the product, it indicates what's the worst or best that could happen. This paper presents a decentralized approach which avoids the central authorities by using blockchain technology. Survey has been done by studying different approaches for making a decentralized system for taking reviews.

Keywords: Blockchain, Smart contracts, Ethereum ,IPFS, Online Consumer Reviews.

I. INTRODUCTION

Blockchain is a distributed ledger technology (DLT), that makes the history of any digital asset unalterable and transparent with use of decentralization and cryptographic hashing. Basically, it is a decentralized, distributed and a digital ledger having records known as blocks. The distributed ledger records the provenance of a digital asset. Each block has cryptographic hash of previous block, connecting the two. The linked blocks form a chain.

An Enormous usage and growth has been seen in online reviews recently. These ratings and reviews are used for various purposes by Tripadvisor, Airbnb, Amazon, Alibaba, eBay etc. The trustworthiness of these reviews has been (and is being) questionable, as these reviews can be tampered by the central authority. Customers give reviews to express their experience with the product. A highly rated product is the first thing the customer look at and can be deciding factor to choose between products [1,2].

Consumers that rely on reviews have to trust atleast two parties involved, i.e the authors of the reviews and the operators of online portals. Untrustworthy reviews of single authors, e.g., an extremely positive review in between negative reviews, can possibly be recognized by consumers. The operators acts as central authority and can modify the data throughout this complete review process. In worst case, they can even exclude consumers from submitting their review [2].

Revain[5,1] is a platform for user reviews that uses IBM's artificial intelligence. It filters out the potential

fake and low quality reviews and the good quality reviews are stored on Ethereum blockchain. The system consists of three components – firstly a stable token system using token (R and RVN). Secondly, IBM 's AI system for filtering out the quality reviews. Finally a immutable blockchain to store reviews so that they never get tampered with.

In this paper, we study a decentralized review system using blockchain to resolve the problem of central authorities. Consumers no longer need to rely on central authorities as blockchains operate decentrally across a network of several nodes, in which every user can participate [2]. Rating and reviews can be stored on the block using smart contracts and assures distributed trust since it is nearly impossible to modify any transaction stored in the blockchain. The information stored in the blockchain is public , i.e anyone can search the chain for historical transactions, enabling consumers to submit and access trustworthy reviews.

Ethereum is a blockchain, developed in 2014 by Vitalik Buterin. In comparison to the Bitcoin blockchain, which handles accounts and transactions, Ethereum stores programming logic also. When paying for its execution, any turing-complete script can be run on Ethereum. Thereby, it enables decentralized apps without any possibility of downtime, censorship, or third-party interference[2,9]. Smart Contracts were first introduced by Szabo [8] in 1994. These are self-executing contracts that operate autonomously. In Ethereum, smart contracts can be written by Solidity. Contracts are executed on several nodes within Ethereum virtual machines (EVM). After executing a contract, nodes must reach a consensus of the calculated result [2,8].Once the consumer gives the review , the review will be stored on to the block and nobody can modify those reviews. These reviews can be referred as untampered reviews [2].Some requirements of this chain –

1) Customer must have purchased the product.
2) Customer must be distinguishable.

After publishing a smart contract, the contract is executed by sending transaction to the contact address. This implementation can be done in two ways.

The first way is by handling the transaction at the backend. The key pair(private and public keys) will be generated at the backend. The signing of transactions also is also done in the backend[2]. The other way is using an app for creation and generation of the keys like metamask.

The drawback of the first way is that the keys are stored on the backend which acts as a central database which can be exploited.

Metamask can be used for generating of keys and helping in transaction. Along with Metamask, customer reviews are stored into the blockchain which can increase the storage cost. To reduce this cost, a distributed file-system can be used such as InterPlanetary File System(IPFS), each item on IPFS can be identified by a hash and can be accessed by URL. The hash came from IPFS, then can be stored onto the blockchain which reduces the storage cost.

II. LITERATURE REVIEW

[1] K. Salah, A. Alfalasi, M. Alfalasi (2019). In this paper they have proposed a blockchain-based review system which is used to store the reviews on the blockchain. They have combined the use of Ethereum networks and smart contracts with Inter-Planetary File System(IPFS). IPFS will act as storage system in which reviews will be stored and the hash of that file will be stored in the

blockchain along with the reviewer's Ethereum address. The Service Provider will issue tokens and the user can submit the review using that token, the token will get verified and the user will be rewarded.

[2] Daniel Martens, Walid Maalej (2018). They have proposed a decentralized review approach which resolves the problem of central authorities, being able to influence the review processes and have also summarized the implementation, challenges and discuss design alternatives.

[3] Ching-Nung Yang, Yi-Cheng Chen, Shih-Yu Chen, Song-Yu Wu (2019). They have introduced a system that tried to resolve a problem of online shopping scams by making an alliance with e-commerce companies and trusted organization to establish a credible product grading system. After verification and product grading , the manager of alliance will write the data onto the blockchain operating a smart contract and the customer may buy product with high grades to insure product quality.

[4] Zibin Zheng,Shaoan Xie, Hong-Ning Dai, Weili Chen, Xiangping Chen, Jian Weng, Muhammad Imran(2019). They have given an overview of blockchain and smart contracts. They have also compared the advantages of smart contract over conventional contracts which includes reducing risks, cutting down administration and service costs and improving the efficiency of business processes. The paper also tells about the creation, execution and deployment of smart contracts and also about creation challenges.

[5] Revain: White paper. This document describes the Revain-plaform. It creates an easy-to-use service that would help companies to receive constructive feedback for their projects. Revain is the first review platform that implements blockchain technology for ensuring the feedback to be genuine and legitimate. In addition, it also discusses about the problems occurring in the current scenario and mathematical formulas for calculating bonuses that make submitting fake reviews unprofitable.

[6] ChengjunCai, HuayiDuan, and Cong Wang, (2018). In this paper, they have establish a set of design principles which will facilitate the development of secure blockchain applications and security management against known attacks like DDoS, Eclipse attacks in the network layer and recovering strategies to deal with massive attacks, consensus protocols and security enhancement.

[7] Gábor Magyar (2017), defines blockchain as the chain of blocks using immutable key referencing mechanism. There is no central authority responsible for governing and managing the data in the blockchain. The values recorded in the blockchain are synchronized between all the peers on the network and the consensus mechanism provides the commonly accepted validity of data. The paper introduces a new approach to an healthcare information model using blockchain.

[8] Nick Szabo. 1994. This paper describes about digital cash protocols as example for smart contracts, defining smart contract as a computerized transaction protocol that executes the terms of a contract , it gives a brief about cryptographic protocols and how smart contracts can bring a difference in the future.

[9] Vitalik Buterin. In this paper author has discussed about the history of bitcoin as a state transition system, mining, merkle trees, about Ethereum and applications of Ethereum.

[10] Daniele Magazzeni and Peter McBurney, William Nash (2017). In this paper, they have explained about blockchain as a type of distributed ledger in which data is stored in

blocks which are linked together to form a sequential chain. They have also discussed about distributed ledger-technologies, shared state, smart contracts and verification and validation of those smart contracts.

[11] Weiqi Dai, Chunkai Dai, Kim-Kwang Raymond Choo, Changze Cui, Deiqing Zou, and Hai Jin (2019). They explain the conventional data trading system which comprises of 3 parties including middle person. The seller will send the dataton the trusted platform and set a selling price. The buyer will select product of interest and order similar to e-commerce transaction. However if the dataset cannot be directly used by buyer, then buyer will need to re-process the dataset to require the results satisfying the buyer's need. This paper analyzes the market and sets new rule for trading in the market using smart contracts to implement secure data trading platform(SDTP) on Ethereum.

III.CONCLUSION

Product's rating highly influence customer's decision. A highly rated product is the first thing the customer looks at and if that rating/review is tampered or manipulated , the customer can get trap into these scams. This paper studied the existed problem which involves central authorities acting as operators. This paper presents a decentralized approach which avoid central authorities by using blockchain technology, enabling customers to submit and retrieve unmodified reviews. This paper also highlight some of the implementation challenges encountered undergoing our approach on the public Ethereum blockchain and discuss possible alternatives.

IV. REFERENCES

[1]. K. Salah, A. Alfalasi, M. Alfalasi (2019). A Blockchain-based System for Online Customer Reviews. 2019 IEEE INFOCOM WKSHPS: CryBlock 2019: Cryptocurrencies and Blockchains for Distributed Systems.

[2]. Daniel Martens, Walid Maalej (2018). ReviewChain: Untampered Product Reviews on the Blockchain. arXiv:1803.01661v1 [cs.CY] 5 Mar 2018.

[3]. Ching-Nung Yang, Yi-Cheng Chen, Shih-Yu Chen, Song-Yu Wu (2019). A Reliable E-commerce Business Model Using Blockchain Based Product Grading System. 2019 the 4th IEEE International Conference on Big Data Analytics.

[4]. Zibin Zheng,Shaoan Xie, Hong-Ning Dai, Weili Chen, Xiangping Chen, Jian Weng, Muhammad Imran(2019). An Overview on Smart Contracts: Challenges, Advances and Platforms.

[5]. Revain: New generation feedback platform based on the blockchaintechnology", Revain Whitepaper. [Online]. .Available: revain.org/pdf/wp/en-wp.pdf. [Accessed: 22-Apr- 2018].

[6]. ChengjunCai, HuayiDuan, and Cong Wang," Tutorial: Building Secure and Trustworthy Blockchain Applications". 2018 IEEE Secure Development Conference 2018

[7]. Gábor Magyar," Blockchain: solving the privacy and research availability tradeoff for EHR data" IEEE 30th Jubilee Neumann Colloquium • November 24-25, •Budapest, Hungary 2017

[8]. Nick Szabo. 1994. Smart contracts. (1994). http://szabo.best.v.wh.net/smart.contracts.html.

[9]. Vitalik Buterin, Ethereum White Paper, A Next Generation Smart Contract & Decentralized Application Platform

[10]. Daniele Magazzeni and Peter McBurney, William Nash (2017). Validation and Verficiation of Smart Contracts: A Research Agenda. Published by the ieee computer society.

[11]. Weiqi Dai, Chunkai Dai, Kim-Kwang Raymond Choo, Changze Cui, Deiqing Zou, and Hai Jin," SDTE: A Secure Blockchain-based Data Trading Ecosystem" IEEE permission. http://www.ieee.org/publications_standards/publications/rights/index.html for more information 2019.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Survey on IOT Based Application for monitoring and Predicting Air Quality in the environment

Vaishnavi Kulkarni[1], Abhijit Kumar Mandal[1], Priyanka Mudaliar[1], Angad Singh[1], Prof. Kranti Kamble[2]
[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India
[2]Professor, Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

India has struggled with pollution for a long time. In fact, in February, India was home to six out of 10 of the world's most polluted cities. In contrast to atmospheric pollution, surrounding pollutants are about 1000 times more likely to be transmitted to the lungs, causing diseases. The main sources of air pollution are motor vehicle emissions, illegal industrial activities, harmful pesticides, and many times we see LPG gas leakages and cylinder truck accidents on road which are harmful for us and can take our lives. Thus, poor air quality causes several health hazards like heart disease, lung cancer, and respiratory problems. The need of the hour is not only to control air pollution but also materializing technologies, devices and software systems to keep a close check on air pollution. Our efforts in this project is to develop an application from which air quality monitoring can be done to take preventative measures to make our living environment safe. The application is user-friendly and works as a preventative mechanism to generate real-time alerts on air quality. Our main contribution is to develop air quality monitoring system that senses the real-time data of surrounding parameters like carbon monoxide, and PM level and alerts the people when the quantity of these elements goes beyond a certain limit and shows the data in an easily understandable format.

Keywords: Air Quality , Raspberry pi, Internet Of Things, Sensors, Air Pollution.

I. INTRODUCTION

Air pollution is one of the environmental issues that cannot be ignored. With deteriorating air quality all over the globe due to industrial and vehicular pollution, there is an increasing risk of breathing problems and lung diseases arising from it. This system which is an android app, is used to sense the air quality of the environment and shows the real condition of air. The purpose of the project is to monitor air quality using different sensors like MQ-135 is an air quality sensor, which is sensitive for benzene, smoke and other pollutant, MQ-4 is a methane and CNG sensor, MQ-7 is carbon monoxide sensor and MQ-6 is a LPG sensor. These are used in combination with the raspberry pi to calculate AQI and notify the user. Solving the drawbacks of existing air quality sensors this application can be used to monitor various gases at a time. The major motivation behind the study and the development of the system is to help the government to devise an indexing system to categories air pollution in India. The project

is to build an air pollution monitoring system application, it's a detection system for multiple information of environment is designed in this project.

II. LITERATURE REVIEW

Ajitesh Kumar, Mona Kumari and Harsh Gupta[1] proposed a cost-proficient air quality observing framework that detects the continuous information of encompassing different boundaries like smoke, carbon monoxide and PM level and cautions the individuals when the amount of these components goes past specific cutoff and shows the information in a without any problem reasonable configuration. Later on, additionally detecting hubs can be added to expand the framework. The significant preferred position of this framework is that it is versatile, little and cost-effective. We have also introduced the execution of an ease IoT based air quality checking framework.

Pratishtha Agnihotri, Sonam Tiwari and Dr. Devendra Mohan[2] proposed a novel scheme to conduct the fine- grained and real-time prediction of AQI based on asynchronous data collected by our monitoring system. They present the asynchronous sensing data and the spatial-temporal-meteorological relations. Based on the CG model, the prediction procedures are carefully designed and an optimization problem arises. They aim to solve the optimization problem by an algorithm combining a closed-form derivation and genetic algorithm. The advantage of the proposed solution over existing ones is evaluated over the data set collected by our monitoring system.

Ravi Kishore Kodali and Sairi pathuri Sasweth C. Rajnarayanan[3] proposed a surrounding air contamination detecting framework can give ongoing estimation of five generally significant for human well- being air boundaries and move it to more significant level applications for examination and anticipating. Estimated information is reinforced along with timestamp and GPS position. The gadget spares information into on-board SD card with capacity to be moved to a have PC by direct USB association or through Wi-Fi transmission.

Zixuan Bai[4] have stated that because of the disturbing degrees of contamination in a portion of the significant urban communities of world, persistent observing of air quality has become a significant issue. By utilizing remote detecting system alongside rapid web association, the checking and examination process has become effective. Such mechanized frameworks are precise, history of the boundaries is moreover put away which can be utilized. The WSN checking frameworks can likewise be executed for different sorts of contamination, water contamination, soil contamination or radioactive pollution.

Mykhailo Lobur and Dmytro Korpyljov and Nazariy Jaworski[5] have presented a remote sensor organize for air contamination observing in light of IOT is useful for the business as well as everyday citizens too. As the contamination information will be accessible with individual and one can see where the contamination level is more around then and the individuals having respiratory sicknesses may abstain from following that way for that specific time of time. Advanced cells are extremely regular now daily. Indeed, even the contamination because of ventures can be checked and the information can be made accessible on the web so that, the important activity to diminish the contamination might be started furthermore.

Md. Mohiuddin Ahmed, Suraiya Banu and Bijan Paul[6] has stated that the alarming levels of pollution in some of the major cities of world, continuous

monitoring of air quality has become a major issue. By using wireless sensing network along with high speed internet connection, the monitoring and analysis process has become effective, efficient and is easily accessible by common people. Such automated systems are accurate, reliable. The WSN monitoring systems can be implemented for water pollution, soil pollution or radioactive contamination. Using high efficiency communication protocol along with sensor network to achieve accurate data in real time.

Octavian A. Postolache, J. M. Dias Pereira, and P. M. B. Silva Girão[7] elaborates the development of an IoT- based indoor air quality monitoring platform is presented. Experiments were performed to verify the air quality measurement device used in the platform based a method suggested by the Ministry of Environment, Korea. We verified the accuracy of indoor air quality monitoring and the desirable performance of the device. Also, experiments making use of the platform were conducted and demonstrated suitable performance and convenience of the air quality monitoring platform. In this paper, the author focused on testing the reliability of the device and implementing the platform.

Somansh Kumar and Ashish Jasuja[8] states that the proposed framework gives minimal effort, low force, conservative and exceptionally exact framework for checking the condition with the committed sensors remotely from any place in this world. An ideal trade off among exactness and cost is accomplished by utilizing single board minicomputer Raspberry pi and proper sensors prompting a well-grounded framework. Air quality observing framework can be more worthwhile if poisons like Sulfur dioxide, nitrogen dioxide, ground level ozone and so forth are additionally checked.

Ajay Chaturvedi and Laxmi Shrivastava[9] has stated the execution of an estimating framework for air quality checking. Two structures are ace postured for remote correspondence between the detecting hubs what's more, a PC that deals with the entire framework. The frameworks are especially appropriate for indoor applications. The yield of the pre-owned gas sensors depends not just on the cross impact of the essential estimated gas yet in addition on external impact factors, to be specific temperature and mugginess. The impact of this cross effect on the exactness of the estimation can be limited utilizing moreover neural systems.

JunHo Jo, ByungWan Jo, JungHoon Kim, SungJun Kim and WoonYong Han[10] states, with the correct usage of the proposed framework they can decrease risky mishaps that happens all through the nation. This arrangement can be introduced in house for house well-being just as any industry or crowed work environment to keep up the air quality safe and lift their work speed. One of the primary explanation of this is outrageous pneumatic force. So, in future they need to expand this framework to recognize pneumatic force of evaporator so it can forestall mishaps and spare numerous lives just as modern misfortune

III.CONCLUSION

Overall, the application is a need of this hour as we know pollution in the environment is increasing at very large huge rate. Due to increase in number of vehicles, the possibility of road accidents has also increased, in which gas trucks are included. By using this application, the users will not only get an alert message prior to explosive action but also, they will receive a proper guidance in reaching to the safe destination. Not only in serious situations, this application also notifies the user with their current

status and gives them tips on how to be more on the safer side.

In future, this application can be extended to the iphone users as currently its an android based application. Also, as per the increase to requirement it can also be extended as website.

IV. REFERENCES

[1]. Ajitesh Kumar, Mona Kumari and Harsh Gupta "Design and Analysis of IoT based air quality monitoring system" Feb 28-29, 2020

[2]. Pratishtha Agnihotri, Sonam Tiwari and Dr. Devendra Mohan" Design of Air pollution monitoring system using wireless sensor network" 2019

[3]. Ravi Kishore Kodali and Sairi pathuri Sasweth C. Rajnarayanan "Smart Indoor air pollution monitoring station" Jan 22-24 2020, Coimbatore India

[4]. Zixuan Bai "Real time prediction for Fine-grained air quality monitoring system with asynchronous sensing", Perking university 2019.

[5]. Mykhailo Lobur and Dmytro Korpyljov and Nazariy Jaworski "Arduino Based Ambient Air Pollution Sensing System" University of Exeter, 2020.

[6]. Md. Mohiuddin Ahmed, Suraiya Banu and Bijan Paul"Real-time Air Quality Monitoring System for Bangladesh's perspective based on Internet of things " International Conference on Electrical Information and Communication Technology (EICT), 7-9 December 2017

[7]. Octavian A. Postolache, J. M. Dias Pereira, and P.M. B. Silva Girão" Smart Sensors Network for Air Quality Monitoring Applications "September 2009.

[8]. Somansh Kumar and Ashish Jasuja "Air Quality Monitoring System Based on IoT using Raspberry Pi " International Conference on Computing, Communication and Automation (ICCCA2017)

[9]. Ajay Chaturvedi and Laxmi Shrivastava "IOT Based Wireless Sensor Network for Air Pollution Monitoring "Jan 2020.

[10]. JunHo Jo, ByungWan Jo, JungHoon Kim, SungJun Kim and WoonYong Han "Development of an IoT- Based Indoor Air Quality Monitoring Platform" January 2020.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

A Survey on online applications for Household Services

Aaditya Pardeshi[1], Kunal Bhalgat[1], Rajeshri Mayanaikar[1], Sayali Desai[1], Prof. Bhagyashree Dhakulkar[2]

[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

[2]Professor, Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

The household service system is incredibly useful for everybody who wants home services like painter, electrician, carpenter, plumber, construction worker, fabricator etc. When an individual relocating from one area to a different one it's hard to find services, they want. In such situation e-Commerce plays an important role in today's life as it has so many advantages in our life because it makes in daily life of the people convenient. So, basically our aim is to design and develop a system that provides many services at your doorstep in just one click. It's a system that helps you connect to one of the best variety of labors providing huge range of services like plumbers, movers and packers, repair persons, cleaners, electricians, painters and many more. The labor will be able to create their role according to their profession and the customer can contact them according to their preferences. There will be no involvement of middle-men in this application. A very simple process is carried out to call a labor for their service. System is versatile as profile of the labors can be viewed from anywhere you desire.

Keywords: Servers [C.5.5], Data Storage and Representation [E.2], User Interfaces [H.5.2], Social Issues [K.4.2], Object-oriented Programming [D.1.5].

I. INTRODUCTION

An online household service system is an android application that basically helps the common man i.e. the customer to connect to our daily wage labour. When someone requires assistance for domestic tasks, the problem occurs due to inaccessibility of service skilled or a trustworthy provider who provides faultless service on request. Most online system for household services provides the most expedient and hassle-free way to get your domestic work done. We aim to help in providing optimal solutions to all your household troubles with more efficiency, ease and majorly, a delicate touch. A single click system describes booking highly skilled in-house professionals and gets your service done on time. This technique helps in providing finest results to all or any domestic troubles with high efficacy and ease. On demand home service system aids not only the users but also the service providers to succeed in out the potential customers. The only main intention of creating this amazing application is to provide the daily wages worker a platform to get into the market and earn for their living, as this Covid_19 pandemic has caused a wave of unemployment amongst them. This online system unlike others can

also be made available in the regional languages as well so that no matter where the user is from, he/she can be able to easily use it. This paper discusses about the web home services, several other services provided and therefore the method of ordering and delivery of services. The event of web based online system helps in determining household services and collaborating interface to look the services. The system also acknowledges the confirmation of services chosen by the users.

II. LITERATURE SURVEY

Taein Hwang [2006] Few service providers have their own service platform to provide services to the customers but the small-scale service providers don't have platform to provide services to customers. It was a high time to provide a common platform so that everyone can participate in the market. This was resolved through digital home service delivery and management system (DSM). It shows how home services of the service providers can be delivered to the service user via the DSM system. One of important point to note here is how the service user can be provided with various services through reliable service aggregator and receive a single bill for the subscribed services.

Chang-Xing Qi [2010] The problem of provincial labor and social security services got resolved through the intelligent switching platform. Intelligent switching platform focus on data interactive mode and inter-departmental information. The architecture of the system was designed in such a way that there was a detailed function of the system. The customer service system is a centralized and integrated platform which is built with call centers and other modem means of communication and information technology.

Shurong Wang [2011] The researcher focusses on the problem of migrant labors and enterprises. He thought of coming up with the platform for migrant labors training and enterprise labor supply. This platform consists of migrant labors information, enterprise information, training organization. Which is all controlled by supervisory authorities. Due to this system overall quality is increases, difficult to find the right people is resolved(training), and many social problems were addressed.

Sheetal Bandekar [2016] "Domestic Android Application for Home Services" is a customized Mobile Application which uses Android SDK (Software Development Kit), Eclipse, Java and MySQL for Android Application Development. This provides domestic services to customers such as: electrical services, plumbing services and carpentry services and many more household services. The application uses GPS to fetch the users' location and assigns nearest service provider from his existing location dynamically. Hence, this application seems to be more dynamic, effective and efficient than the existing system.

Zhang Fuyan [2017] Through this paper we come to know how the problem in a household could be resolved using the STM32core in which they designed a robot which has

sensors and Wi-Fi camera loader mounted on the machine to remote control the moments and all so obtain the environmental data through sensors. PHP and QT is used to build the server. Robot and mobile communicate through telecommunication technology, whereas for location and tracing purpose Bluetooth was used.

N. m. Indravasan [2018] noted that in this modern life style, people are very much in their heavy work culture. In the busy schedule if any unexpected household task pops up. That distracts them from their work. E-Commerce can play a primary role in solving this issue. Creating a platform that can provide number of services in one click. Researcher

went with email confirmation for verifying the customers who will be using the platform.

Neale A. Dagdag [2019] This was a completely mobile application for android device. Where the main aim was finding work opportunities for skilled workers. The skilled workers will get connected to customers who need services such as: electrical service, plumbing service, automotive repair, and other similar services which can be done at the customer's home. The main revenue shall come from commissions and quarterly membership fee from the skilled workers. Revenues may also come from advertisers and/or companies who wish to tie-up with At-Your-Service mobile application.

S Rachitha [2019] Focus on establishing a market platform which not only embrace consumer demand but also providing platform for service professionals an opportunity to earn additional income. Researchers came up with an idea of providing an option to the people, wherein if they encounter any issue, they can contact a service professional from another location as well. So, who can assist them in fixing the issue? Web application is deployed on cloud which will reduce the infrastructure overhead and operational cost. System also provides quality of service, as it works on rating and servicemen working history.

K. Aravindhan [2020] This is an online home services system. The purpose was to find the service providers detailed information which helps customers to get their services fulfil instantly. The thing which makes it different from other system is "chatbot" which helps the users to clarify the queries posted.

III.CONCLUSION

This paper lists the analysis of the household service system that helps to reduce the burden in finding appropriate service provider by providing a detailed information which helps the users to get their service fulfil instantly. A systematic android application offers ease in accessing the required services in a more comfortable way. With a well-qualified and background demonstrated professional we may provide all plumbing, carpenter, electrician, fabricator, construction worker and many more services to be done in a click anytime and from anywhere. There is a need to create a more dynamic application rather than the existing system. As technology is reaching heights day by day, we too should make sure we make the best use of it by helping others out and giving Therefore, by the survey performed we conclude that that in today's era booking a service through mobile application will be a new way of communicating and getting your work done. Thus, we can make an effort to take this technology forward and to make the best out of it would be our only goal.

IV. REFERENCES

[1]. K.Aravindhan, K.Periyakaruppam, T.S Anusa, S.Kousika, A.Lakshmi Priya, "Web Application Based On Demand HOme Service System"6th International Conference on Advanced Computing And Communication Systems(ICACCS) ,2020.

[2]. Neale A. Dagdag , Almar Allan F. De Guzman , Rowena V. Pamplega, Grace Lorraine D. Intal, " At-Your-Service Mobile Application: E-Hub for Skilled Workers"IEEE 6th International Conference on Industrial Engineering and Application,2019.

[3]. Nikam Poonam, Gunjal Trupti, Jadhav Priti, Parakhe Sonali, Prachi Tambe, "Survey on Home Provider" International Research Journal of Engineering and Technology (IRJET) December 2019.

[4]. N.M. Indravasan, Adarsh G, Shruthi C, Shanthi K, "An Online System for Household Services" International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, May 2018.

[5]. Zhang Fuyan, Teng Yingyan ,"Design and Realization of Household Service Robot Based on STM32 and Server" International Conference on Robots and Intelligent System, 2017.

[6]. Sheetal Bandekar, Avril D'Silva, "Domestic Android Application for Home Services" International Journal of Computer Applications, ISSN No.0975 – 8887, Volume 148 – No.6, August 2016.

[7]. Shahrzad Shahriari, Mohammadreza Shahriari, Saeid gheiji, " Ecommerce And It Impactson Global Trend And Market" International Journal of Research – Granthaalayah. Vol.3 (Iss.4): April 2015

[8]. Shurong Wang, "The Architecture Design of Migrant Labors Training Employment Information Platform" IEEE 6th International Conference on Industrial Engineering and Application,2011.

[9]. Chang-Xing Qi, Qing-Dong Du,Hong-Wei Wang ,"Construction of Provincial Labor and Social Security Customer Service System" International Conference On Computer Design And Appliations, 2010.

[10]. Taein Hwang, Hojin Park, Jin Wook Chung, "Design and Implementation of the Home Service Delivery and Management System Based on OSGi Platform", IEEE 2006.

[11]. Bhagyashree Dhakulkar Arati.B.Burde, Prachiti.R.Shahare, Priya Desai, "Civic Enhancement Versatile Approach" International Journal of Engineering Technology Science and Research, ISSN 2394 – 3386 Issue 10,2017

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Face and Food Recognition Techniques using Convolutional Neural Network

Nidhi Shah[1], Vanshika Singh[1], Pranjal Kumavat[1], Samiksha Shirbhate[1], Dr. Meenakshi A. Thalor[1]
[1]Information Technology, B.E AISSMS's Institute of Information Technology, Pune, Maharashtra, India

ABSTRACT

There are different face recognition and food recognition techniques that have been studied and implemented in the past. As the world is moving in a fast pace security is an important aspect of a person's life. Apart from that healthy food has its own benefits. In this paper, various techniques for face recognition and food recognition have been listed along with their methodologies and accuracies. This paper proposes a model which will detect face and food from the given input images using various algorithms and then recognition of face and food item will be done.

Keywords: Face recognition, Food recognition, CNN, EigenFaces, Haar

I. INTRODUCTION

Face recognition is used in some of the high-tech security systems. With the use of videos or images and real time face detection helps in maintaining proper reports and also helps in security management. Maintaining a good diet starts with eating healthy food. So, recognizing the food items we are consuming on daily basis is also important. These processes are made automatic using different algorithms on various datasets. In the past, there have been various techniques for face detection to recognition. These techniques have evolved over a significant timeline. We have discussed different techniques like Eigen Faces, LBPH, CNN for face recognition. Similarly, in food recognition different techniques have been discussed. Each paper has used different datasets and has proposed a model with the highest achievable accuracy. In this paper we have presented a survey on different techniques used for face recognition and food recognition separately. Methodology and conclusions of these papers are explained along with the dataset used. Section 2 explores the literature review of these papers. In Section 3, we have discussed our proposed model and finally in the last section includes conclusion and future scope of face and food recognition.

II. LITERATURE SURVEY

A. Face Recognition Techniques

In [1], face detection and recognition are done on real time basis. Videos are taken as input, where frames are selected using an algorithm for further process. The Haar algorithm is used for face detection. This algorithm performs well for illumination changes. Recognition of faces is done using EigenFaces and Gabor Feature algorithms. EigenFaces algorithm performs well in complex computations and Gabor feature performs well when there are pose changes. The combination is proposed as a robust system. The accuracy of the system in recognizing a face using Eigen face is approximately 60%-65% and using Gabor feature is approximately 50%-60%. This system

will work well under limited conditions and dataset. Real time recognition can be achieved using other methods like CNN.

In [2] as well, the Haar method is used for face detection. A real time attendance system is proposed which detects faces of students using different modes like snap and video. It is built on Raspberry Pi and software module uses Open source computer vision libraries. The face detection here is done using the Haar Cascade method which classifies the input images as "with face" and "without face" images. It is a feature-based method. Recognition of the faces is done using LBPH method. Firstly, LBPs are obtained for the image and histograms are plotted which are then compared with the training dataset. Hence face is recognized whatever the input type is provided. The accuracy of this system is 74%. The accuracy has increased due to the use of EigenFaces for recognition. Another attendance system is proposed in [3],

which uses Haar face detection and PCA and EigenFaces for recognition of faces. The proposed system stores the records of the students in the database. The captured image undergoes the detection process. The process is carried out by checking the landmarks of the face using AdaBoost Algorithm. The recognition is done by Principle Component Analysis with EigenFaces.

In [4], Convolutional Neural Network (CNN) is described for face detection. CNN is used for image classification and object recognition as it works on raw pixel strength of image, which is given as input as flat vector. The collected data is filtered, split and labelled for training. Each image is converted to one fixed size and the model is trained. The ReLu activation function, set dropout proportion to 100% and first order optimization algorithm is used to improve the model efficiency. The accuracy of model is increased from 68.85% to 79.41% after the test set.

In [5], CNN technique on FPGA is used for face recognition. Generally, CNN is implemented using CPU and GPU. This paper proposes real time face recognition by using FPGA. The accuracy of the network is determined by the testing unit. The system built on FPGA consists of 6 CNN layers. [5] Fig. 1 shows the CNN layers. There are 4 hidden layers where there are 2 convolutional layers and 2 pooling layer. The activation function determines the sigmoid function. This function determines the error rate of the network. The maximum error seen in the network after training is of 0.001. The system claims to be 99.25% better than other systems built on CPU, GPU and even other algorithms like PCA and LBP. Network is trained and tested on face datasets having a range of different expressions, lights and other factors. The system shows good robustness. The only drawback of this is that there isn't much research done on face recognition using FPGA. It is also difficult to code a CNN algorithm on a FPGA as there is yet no mature research in this field. The hardware of this system can be a limiting factor.

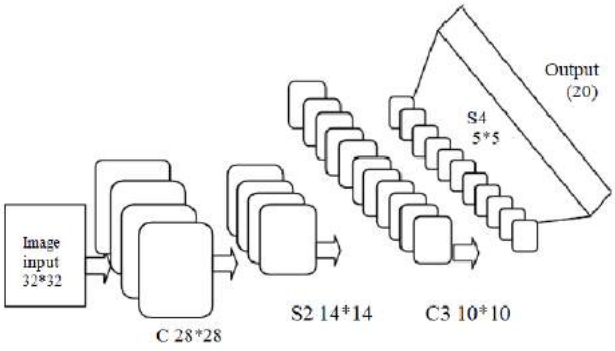


Figure 1: CNN of Face Recognition on FPGA

In [6] specifically focuses on face recognition when there are some obstacles in capturing the face. As in real time getting a clear image of the face and recognition can be difficult. This paper mainly focuses on recognition of face when it is partially visible. CNN is used for extracting different features from an occluded face image. It is done by weighted neurons which shred on a weighted plane. This decreases the computational complexity of the model. The recognition of the face is done by calculating the

Euclidean distance. Euclidean distance is calculated of the input image and the referenced image. It should be less than 1 if it is the same person. The triplet loss function is used in the paper for determining the degree of inconsistency. In this the dataset had 30% occlusion over the faces. The model proposed in this paper has a recognition rate of 98%.

Multiple layer model based real time face detection [7] works on CPU. Different ways are used are used to cover the face for training the model for real-time detection. This system works on CPU using different hardware. Multiple Images are detected using CNN architecture which uses shallow layers that produce light variations so the system can work as close to real- time. This system has built a fast face detection system. They have achieved this using Average Precision for measuring accuracy of the system which is 90.84%.This can be further improved by using pre-processing.

B. **Food Recognition Techniques**

In [8] deep convolution neural network (DCNN) is proposed to classify food images. Food image analysis has three steps: 1. image segmentation 2.food recognition 3.quantity analysis. In food analysis colour, intensity, texture, etc are important features. The model proposed consists of 54 layers.3 datasets have been used for training and testing of the model. In this approach the network is pre- trained using Inception v3.Feature extraction is done using this Inception v3. To improve the performance of this type of network increasing its size, depth layers, width makes a huge difference. The model shows 92.3% accuracy in food detection. The only drawback of this model is that even after using pre trained model the computational resource requirement is quite high.

This paper [9] uses ensemble network containing GoogleNet, ResNet, and Alexnet for food image detection. The model of this CNN has been trained and tested on 2 types of databases. One is the ETH-101 and the other is Indian food database. The feature extraction is done with the help of external pre-trained model. AlexNet and GoogleNet are used in the max pooling layer of the CNN module and ResNet is used in the last output layer. The accuracy of this model is 73% for Indian food database which in comparison to AlexNet, ResNet and GoogleNet individually is quite high. The hardware used in this experiment include NVIDEA processor with system requirement of 128GB of ram. the parameters used in the sub-network is different which adds to the complexity as reshaping the parameters according the sub-network increases load. As the architecture of each sub-network is different, it could be challenge to build such a network. The building of such a networking module could be difficult to code and debugging might be a challenge in future.

The paper [10] proposes a method for food image classification and furthermore how to determine its calories. Food recognition is done by 2 techniques namely, Graph cut segmentation and deep learning neural network. From these two methods CNN has proven to be more efficient comparatively in recognizing the food and hence providing its calorie count. The application in the paper calculates probability of the input image against the labelled data in the database hence predicting the food in the input image. Calorie measurement of the recognized food item is done by 2 methods.

1) Finger Based Calorie Measurement.
2) Calorie Measurement using Distance Estimation.

The application shows more than 95% accuracy in identifying the food and showing perfect calorie count. The database used to train and test the model contained images of only single food items.

The CNN approach proposed in paper [11] is built on a 5-layer network architecture. The first 4 layers are convolutional-pooling layers and last layer is a fully connected layer. The steps in recognizing the food

given in the input image are as follows: Feature Extraction, Shift and distortion invariance and Classification. This Fig. 2 [10] shows the system architecture used. Different layers of the CNN perform these steps and the output is provided. UEC-FOOD100 and an open source database are used in training and testing of model. The accuracy on single food item is 80.8% and 60% on multi- food item datasets.

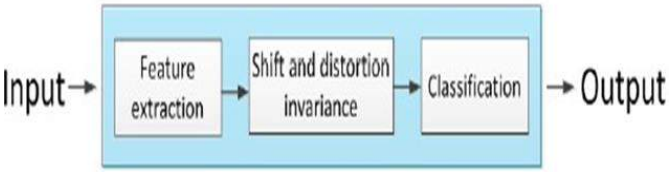


Figure 2: System Architecture

III. PROPOSED MODEL

In reference to above literature survey, CNN has performed well under all the different scenarios. We propose a system which will detect face and food from the given input images using algorithms and further recognize the face and food item. We will be training the system under supervised manner. The feature extraction and further classification will be done by CNN.

Face detection is done using MT-CNN algorithm which detects the faces in the given input image and crops it to 224 * 224 size array. This array is input to extended model of VGG-Face model which performs recognition of the faces and marks attendance of the person accordingly. The VGG- Face model is used after altering its output layer with respect to our dataset. The process of using the pre-trained model as a feature extractor for a new model is called transfer learning. Only the output layer is fine-tuned according to our dataset.

Food Recognition is done using VGG-16 model. This model is mainly used for object detection and classification of objects in a photograph. We will be using this model after fine tuning its output dense layers. We have created our own dataset of purely Indian dishes. Some of the images are taken from the internet and some are taken by us of some dishes which are common in our household. The models will be given input of the images which are uploaded on a website. The final output of these models will also be shown on the website itself. The output will also be stored in the database for future reference. Fig.3 show the simple architecture of proposed system.

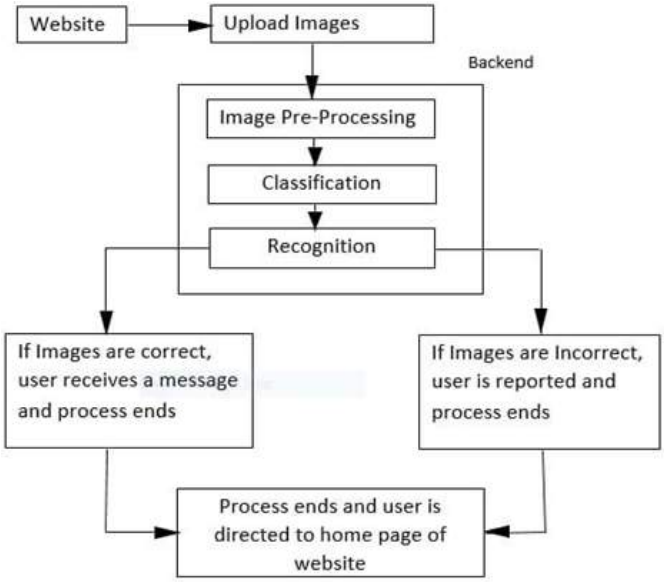


Figure 3: Simple Architecture of Proposed System

IV. CONCLUSION

In this review it is seen that CNN shows the best accuracy among the algorithms, compared for both face and food recognition. It can be implemented on medium database efficiently. It works efficiently despite the occlusion, pose, expression and light change in face recognition. Similarly, for food recognition colour, texture and multi-food items are recognized with high accuracy. As a whole these will help in making the society safer to place to live in and even healthier.

V. REFERENCES

[1]. S. V. Tathe ,A. S. Narote,S. P. Narote, " Face Detection and Recognition in Videos", IEEE. Bangalore Section, Institute of Electrical and Electronics Engineers. India Council, & Institute of Electrical and Electronics Engineers. (n.d.). 2016 IEEE Annual India Conference (INDICON) : 16-18 Dec. 2016.

[2]. V, T., Vellore Institute of Technology, Institute of Electrical and Electronics Engineers. Madras Section, IEEE Communications Society., & Institute of Electrical and Electronics Engineers. (n.d.). Conference proceedings, International Conference on Vision Towards Emerging Trends In Communication and Networking (ViTECoN 2019) : 30-31, March 2019, Vellore, Tamilnadu, India.

[3]. Shreyak Sawhney, Karan Kacker,Samyak Jain, Shailendra Narayan Singh, Rakesh Garg,"Real-Time Smart Attendance System using Face Recognition Techniques" 9th International Conference On Cloud Computing, Data Science and Engineering : Confluence 2019 : 10-11 January 2019, Uttar Pradesh, India.

[4]. Di Wang , Ding Wang , Hongzhi Yu , Guanyu Li " Face Recognition System Based On CNN" 978-1-7281-9837- 8/20/$31.00 ©2020 IEEE

[5]. Qu, X., Wei, T., Peng, C., & Du, P. (2018). A Fast Face Recognition System Based on Deep Learning. Proceedings- 2018 11th International Symposium on Computational Intelligence and Design, ISCID 2018, 1, 289–292. https://doi.org/10.1109/ISCID.2018.00072.

[6]. Gui Wu1, Jun Tao, Xun Xu " Occluded Face Recognition Based on the Deep Learning "IEEE 31st Chinese Control and Decision Conference (2019 CCDC) : 3-5 June, 2019, Nanchang, China.

[7]. M. D. Putro, Wahyono and K. -H. Jo, "Multiple Layered Deep Learning Based Real-time Face Detection," 2019 5th International Conference on Science and Technology (ICST), Yogyakarta, Indonesia, 2019, pp. 1-5. doi: 10.1109/ICST47872.2019.9166172

[8]. Hassannejad, H., Matrella, G., Ciampolini, P., De Munari, I., Mordonini, M., & Cagnoni, S. (2016). Food image recognition using very deep convolutional networks. MADiMa 2016 - Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management, Co-Located with ACM Multimedia 2016, 41–49. https://doi.org/10.1145/2986035.2986042

[9]. Pandey, P., Deepthi, A., Mandal, B., & Puhan, N. B. (2017). FoodNet: Recognizing Foods Using Ensemble of Deep Networks. IEEE Signal Processing Letters, 24(12), 1758– 1762. https://doi.org/10.1109/LSP.2017.2758862

[10]. Pouladzadeh, P., Kuhad, P., Peddi, V. B., Yassine, A., & Shirmohammadi, S. (n.d.). Food Calorie Measurement Using Deep Learning Neural Network.

[11]. Zhang, W., Zhao, D., Gong, W., Li, Z., Lu, Q., & Yang, S. (2015). Food Image Recognition with Convolutional Neural Networks. 2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom), 690–693. https://doi.org/10.1109/UIC-ATC-ScalCom-CBDCom- IoP.2015.139

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

A Review on Cloud Data Security Challenges and Opportunities Related to Big Data

Ms Hema B. Sable[1], Dr. Manoj S. Kathane[2]

[1]Research Scholar, Electronics & Telecommunication Department, MGMCET, Kamothe, Navi Mumbai, Maharashtra, India

[2]Research Guide Principal, TSEC, Burhanpur, Madhya Pradesh, India.

ABSTRACT

Now days in every sector of IT, cloud computing is inveloved and implemented. Cloud computing with big data application is additional reward and provides the important services to the society.Cloud computing offers the plenty of services to the users by using the internet. However the users are victim because of unsecurity and unauthorised access. By expansion in using the network and sharing data between the end users, it is very important to maintain the data security and privacy.But still after the tremendious improvement in the technology to provide security for the data is big challenge infront of IT technology. So it is very prime thing to provide security to the user's data in cloud computing services which are interconnected with big data.Big data provides a great profit to organizations, commerce, companies and many large scale and small scale industries.

Keywords:- Data Security, Cloud computing and Big data

I. INTRODUCTION

While dealing with complex data it is very important to store the data securely. Data sharing should be free from the unauthorised accessed. Cloud has the features that it never gives the control to the user while storing the data on the cloud; user never knows the location where the data is placed on the cloud. The explanation following this control issue is that if one desire to acquire the profit of cloud computing, user must operate to share the resources and also the scheduling given by the controls [2]. It is very important to remove this complexity in the data computing. For this cloud computing is the basic and firm Dias to perform and remove the large scale complexities.The main goal of this cloud computing technology is that to provide solution and different method for handling Big Data [4].

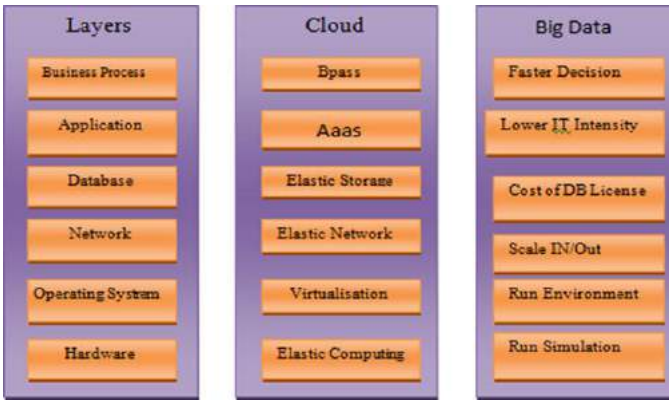


Fig1: Big Data and Cloud [4]

Fig1. Shows the different layers in the cloud and Big Data technology which provide the complexity solution and keep data safe from the untrusted access.

1. Cloud Computing

Initialy, because of huge involvement of the users in accessing the resources, IT sector was failed to provide the required resources and many stakeholders could not complete their projects due to lack of availability of resources.But today many users can access many resources for the unlimited time period as cloud computing provide this facility to the end users.This technology provide paid resources to the users so that they can used any type of resourses (either hardware or software) according to their requirement. Cloud is set of computing devices and nodes that shares services publically on the World Wide Web. Cloud implemented and managed applications by gathering all computing resources together also maintains user's history information in their internet usages, processes and offer accurate information, so that it offers intellectual support service to IT users[1].

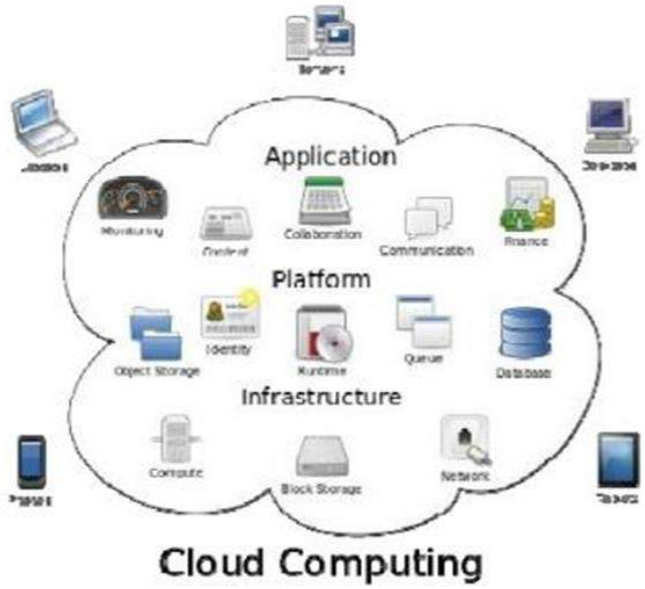


Fig2: Cloud computing environment [9]

2. Big Data

Big data is noting but the high volume of data which is categories into two main category one is structured data and other is unstructured data that is too much large that is very hard to used and process this data by using traditional databases and software technologies. The Big data has the following properties.

a) Volume: Many factors contribute towards increasing Volume streaming data and data collected from sensors etc.
b) Variety: Today data comes in all types of formats emails, video, audio, transactions etc.,
c) Velocity: This means how fast the data is being produced and how fast the data needs to be processed to meet the demand[4,3].

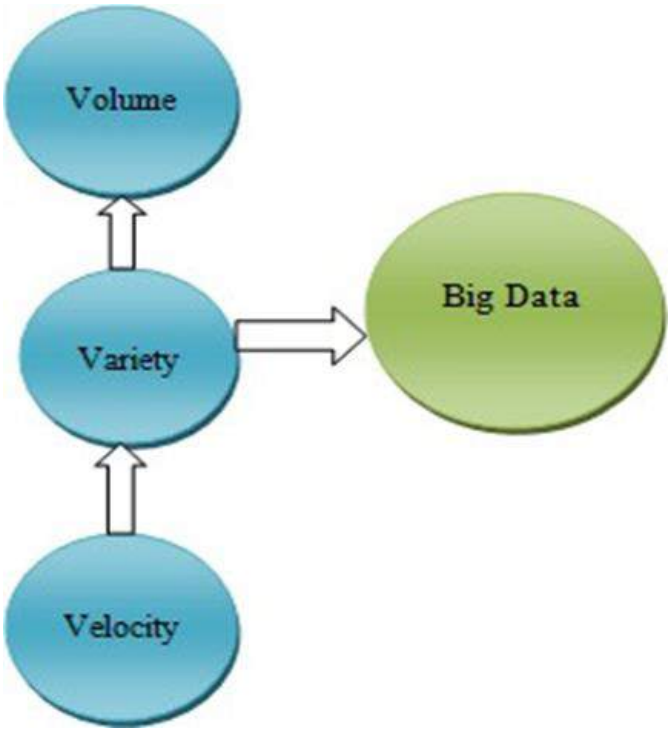


Fig3: V criteria of Big Data[6]

II. CLOUD COMPUTING IN BIG DATA

The increase of cloud computing and cloud data stores has been a forerunner and facilitator to the emergence of massive data.Cloud computing is that the commoditization of computing time and data storage by means of standardized technologies. It has important advantages over traditional physical deployments. However, cloud platforms are available several forms andsometimes need to be integrated with traditional architectures. This results in mystify for decision makers responsible of massive data projects, results in an issue of how and which cloud computing is that the optimal choice for his or her computing needs, especially if it's an enormous data project? These projects frequently show changeable, stuffed, or immense computing power and storage needs. At an equivalent time business stakeholders expect swift, inexpensive, and reliable products and project outcomes[4].

III. BIG DATA APPLICATIONS

The big data application refers to the massive scale distributed applications which usually work with large data sets. Data exploration and analysis became a difficult problemin many sectors within the span of massive data.With large and sophisticated data, computation becomesdifficult to be handled by the traditional processing applications which triggers the event of massive data applications . Google's map reduce framework and apache Hadoop are the defacto software systems for giant data applications, during which these applications generates an enormous amount of intermediate data. Manufacturing and Bioinformatics are the 2 major areas of massive data applications. Big data provide an infrastructure for transparency in manufacturing industry, which has the power to unravel uncertainties like inconsistent component performance and availability. In these big data applications, a conceptual framework of predictive manufacturing begins with data acquisition where there's an opportunity to accumulate differing types of sensory data like pressure, vibration, acoustics, voltage, current, and controller data. The combination of sensory data and historical data constructs the big data in anufacturing. This generated big data from the above combination acts because the input into predictive tools and preventive strategies like prognostics and health management. Another important application for Hadoop is Bioinformatics which covers subsequent generation sequencing and other biological domains. Bioinformatics which requires a large scale data analysis, uses Hadoop. Cloud computing gets the parallel distributed computingframework alongside computer clusters and web interfaces[4].

IV. ADVANTAGES OF BIGDATA

Big data, the software packages provide an upscale set of tools and options where a private could map the whole data landscape across the corporate , thus allowing the individual to analyze the threats he/she faces internally. There are some common characteristics of massive data, like a) Big data integrates both structured and unstructured data. b) Addresses speed and scalability, mobility and security, flexibility and stability. c) In big data the belief time to information is critical to extract value from various data sources, including mobile devices, frequency identification, the online and a growing list of automated sensory technologies. All the organizations and business would enjoy speed, capacity, and scalability of cloud storage. Moreover, end users can visualize the info and corporations can find new business opportunities.If big data are combined with predictive analytics, it produces a challenge for several industries. The combination

leads to the exploration of those four areas: a) Calculate the risks on large portfolios b) Detect, prevent, and re-audit financial fraud c) Improve delinquent collections d) Execute high value marketing campaigns[5]

V. NEED OF SECURITY IN BIGDATA

IT Vendors offers storage, computation, application hosting services with backed performance and charge on pay-per- use techniques via cloud service providers (CSP). But also, cloud computing undergoes several security issues and defined below: i) Data occurrences of user data has got to be integrated, linked, mapped and transformed from other occurrences of another user data. ii) Data migration over the web exposes sensitive information to malicious intrudes results in loss of privacy, legal liability and reputational loss to the IT organizations. iii) Internet is hospitable global users, an intruder can exploit a software bug to smell trustworthy information, to misuse IT resources, or generates a pathway resulting in advanced threats and attacks. iv) Outsourcing information must be limited to a authorized users with access rights, so to avoid the abuse of malicious access. v) Lack of knowledge replication and recovery technique in cloud computing results in loss of knowledge .vi) Cloud services must be made accountable mandatory as an application requirement[1].

VI. ISSUES AND CHALLENGES

Cloud computing comes with numerous security issues because it encompasses many technologies including networks, databases, operating systems, virtualization, resource scheduling, transaction management, load balancing, concurrency control and memory management. Hence, security problems with these systems and technologies are applicable to cloud computing. For example,it is vital for the network which interconnects the systems during a cloud to be secure. Also, virtualization paradigm in cloud computing leads to several security concerns. For example, mapping of the virtual machines to the physical machines has got to be performed very securely. Data security not only involves the encryption of the info , but also ensures that appropriate policies are enforced for data sharing. In addition, resource allocation and memory management algorithms even have to be secure. The big data issues are most acutely felt in certain industries, like telecoms, web marketing and advertising, retail and financial services, and certain government activities. The data explosion goes to form life difficult in many industries, and therefore the companies will gain considerable advantage which is capable to adapt well and gain the power to analyze such data explosions over those other companies. Finally, data processing techniques are often utilized in the malware detection in clouds. The challengesof security in cloud computing environments is often categorized into network level, user authentication level, data level, and genericissues. Network level: The challenges which will be categorized under a network level affect network protocols and network security, like distributed nodes, distributed data, Internode communication. Authentication level: The challenges which will be categorized under user authentication level deals with encryption/decryption techniques, authentication methods such asadministrative rights for nodes, authentication of applications and nodes, andlogging. Data level :The challenges that can be categorized under data level deals with data integrity an availability such as data protection and distributed data. Generic types: The challenges which will be categorized under general level are traditional

security tools, anduse of various technologies [4,6,7,8].

VII. CONCLUSION

Cloud surroundings are widely utilized in industry and research aspects; therefore security is an important feature for organizations management on these cloud environments. Using proposed approaches, cloud environments are often secured for complex business operations.

VIII. ACKNOWLEDGEMENTS

The authors would like to thank the editor, the guest editors and all reviewers for valuable comments on an earlier version of the manuscript.

IX. REFERENCES

[1]. DR.S.P.ANANDARAJ, MR.MOHAMMED KEMAL," RESEARCH OPPORTUNITIES AND CHALLENGES OF SECURITY CONCERNS ASSOCIATED WITH BIG DATA IN CLOUD COMPUTING", INTERNATIONAL CONFERENCE ON I-SMAC (IOT IN SOCIAL, MOBILE, ANALYTICS AND CLOUD) (I-SMAC 2017)

[2]. Venkata Narasimha Inukollu, Sailaja Arsi and Srinivasa Rao Ravuri," Security Issues Associated With Big Data In Cloud Computing", International Journal of Network Security & Its Applications (IJNSA),Vol.6, No.3, May 2014

[3]. Saeed Banaeian Fa, Azadeh Imani Rad," Security Analysis of Big Data on Internet of Things", arXiv: 1808.09491v1 [cs.CR] 28 Aug 2018

[4]. K.Shanmugapriya, M.Murugeswari, K.Suriya," Security Issues Associated With Big Data in Cloud Computing", K.Shanmugapriya et al, /(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (6) , 2015, 4952-4956.

[5]. R.Saranya, V.P.MuthuKumar," Security issues associated with big data in cloud computing", International Journal of Multidisciplinary Research and Development, Volume: 2, Issue: 4, 580-585 April 2015 www.allsubjectjournal.com e-ISSN: 2349-4182 p-ISSN: 2349-5979 Impact Factor: 3.762.

[6]. Bala M. Balachandran and Shivika Prasad," Challenges and Benefits of Deploying Big Data Analytics in the Cloud for Business Intelligence", International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2017, 6-8 September 2017, Marseille, France.

[7]. Qusay Kanaan Kadhim , Robiah Yusof , Hamid Sadeq Mahdi , Sayed Samer Ali Al-shami , Siti Rahayu Selamat," A Review Study on Cloud Computing Issues", 1st International Conference on Big Data and Cloud Computing (ICoBiC) 2017 IOP Publishing IOP Conf. Series: Journal of Physics: Conf. Series 1018 (2018) 012006 doi :10.1088/1742- 6596/1018/1/012006.

[8]. Sitalakshmi Venkatraman and Ramanathan Venkatraman," Review Big data security challenges and strategies", AIMS Mathematics, 4(3): 860– 879. DOI: 10.3934/math.2019.3.860 Received: 25 April 2019 Accepted: 01 July 2019 Published: 19 July 2019.

[9]. Sheetal Singh, Vipin Kumar Rathi, Bhawna Chaudhary," Big Data and Cloud Computing: Challenges and Opportunities", International Journal of Innovations in Engineering and Technology (IJIET)

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Significance of Internet of Things with MQTT Protocol for Light Weight Communication

Deepali S Hirolikar[1], Smita A Kakade[1], Dr. Pankaj Agarkar[2]
[1]Department of Information Technology, Shri JJT University, Jhunjhunu, Rajasthan, India
[2]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

The machine to machine interaction unit shows several or perhaps even more systems that may immediately hook up as well as , correspond amongst one another, alternatively they may using an intermediary utility server. These types of units converse through various categories of channels, consisting of IP systems or maybe the Cyberspace. This paper presents the MQTT protocol commonly utilized for IoT communication. Publish-Subscribe are actually alluded to a communicating unit which usually consists of publishers, brokers as well as, consumers. Publishers will be the resource of information and facts. Publishers post the information and facts to the topics that are monitored by the broker.

Keywords: IoT, MQTT, Cyberspace

I. INTRODUCTION

An MQTT [1,2,3] session is usually divided into four phases: interconnection, authentication, end of contract and interaction. A client begins by constructing a Transmitting Control Standard protocol/Internet Protocol (TCP/IP) connection to the broker by employing either a regular interface or a custom made dock described by the broker's providers. When formulating the interconnection, it is definitely essential to identify that the hardware might continue an aged program if it can be offered with a used again customer identification. The regular ports will be 1883 meant for nonencrypted communication and 8883 for protected conversation using SSL or (TL). At the time of many of these handshake, the consumer validates the server qualification as well as , authenticates the machine [4,5]. The client can likewise offer a customer record to the broker within the handshake. The broker may make use of this to authenticate the customer. Even though not really particularly component of the MQTT standards, it offers turn into normal for agents to assist consumer authentication with SSL/TLS client-side accreditation

As the MQTT protocol is designed to get a process for source confined and IoT products, SSL/TLS [6,7] may certainly not usually come to be a choice as well as, in several situations, might not be preferred. On such events, authentication is certainly offered as an obvious text message username and security password, which will be delivered by means of the customer to the machine this, as component of the CONNECT/CONNACK box series [8].

In addition, some broker agents specifically open up brokerages released on the net, will acknowledge

private customers. In many of these circumstances, the user name as well as , password will be just remaining empty. MQTT is normally known as a light-weight process as all its communications possess a little code impact. Every concept includes a set header 2 bytes an various adjustable header, a concept payload that is reserved to 256 MB of info and so a QoS level [9,10]. During the conversation stage, a customer can carry out publish, subscribe, ping operations and unsubscribe. The publish procedure transmits a binary stop of info to a subject that is usually described by way of the author.

II. M LITERATURE REVIEW

Creator provided a study of IoT and Cloud Work because of a concentrate on the security problems of simultaneously technology. Particularly, author mixed the two above mentioned solutions in buy to analyse the prevalent aspects, as well as , in order to understand the advantages of their integration. Finishing, writer presented the impact of Cloud Research to the IoT concept. Therefore, it demonstrated just how the Cloud Computing technology, increased the action of the IoT. At last, creator targeted the reliability problems of the incorporation of IoT and so Cloud Computing

Individuals, health professionals, medical companies, and experts be based upon analytical versions produced from many of these info resources to through the network keep an eye on individuals, early-diagnose illnesses, and discovered customized remedies as well as , medicines. Nevertheless, with no suitable personal privacy safety, performing data statistics turns into a resource of a personal privacy headache. In this kind of content, publisher offered the study issues in producing useful level of privacy conserving analytics in medical care info devices. Publisher assessed the info and then analytic preferences for the included celebrations, determine the privateness property, reviewed known privacy substrates, and so talked about the feasible trade-off among personal privacy, effectiveness, and unit top quality [11].

Catastrophe supervision required a close to current data diffusion to ensure that the crisis solutions can become supplied to the proper persons at the right period. Latest improvements in facts and conversation systems allow group of current information from numerous resources. Writer also, presented this sort of a scenario conscious program to assist catastrophe control data lifecycle, i.e. from info intake and control to notify scattering [12].

Wise Golf course Residence android application was being successful to notice and controlling the microclimatic ground inside a property. Right from the renewable home very easily obtain ground dampness, moisture and heat sensor worth to android software , relating to detectors beliefs and then publisher arranged predetermined tolerance principles for every sensor ,based on sensor psychic readings author had been heading to restrain employing drinking water sprayer , chilling lover , roof and concentrate light and so simply press the switch in android author can help to make on/off engines and it likewise experienced datasheet of all garden planting as well as , time of year smart safety measure materials for monitoring and managing [13].

The period of program integration with the aid of service-oriented architecture is usually quickly getting a heritage. New systems progressively depend on the heir tiny support structures. Nevertheless, allocated uses can easily as well perform a critical part in the internet-of-things. In this newspaper, writer collected proof showing cross-cutting problems in these kinds of areas as well as, spotlight how aspect-oriented encoding details them. Particularly, the goal of this paper is usually to offer a roadmap meant for using aspect-oriented coding to offer efficiently by

way of crosscuts in request style, devices conversation, and incorporation [14].

The Internet of Things (IoT) is certainly an important element of Industry 4.0. Scheduled to developing requirements of clients, the current IoT design will not really end up being dependable and reactive for following era IoT applications as well as, forthcoming companies. Info Analytics and Big Data are being used meant for recognition of overall performance. This mixture of technologies is usually in a position to meet requirements of fresh applications. The proposed book architecture is usually modular, effective, agile, scalable, basic, and so it is normally ready to gratify the large figure of data and program demands [15].

With the developments in Net systems as well as, Cellular Sensor Systems, a fresh period of the Internet of Things (IoT) is becoming recognized. IoT created a great deal of details which can be utilized to enhance the performance of their daily world and gives advanced products and services in an array of request domain names. Therefore, a privacy-protected and inter-cloud data fusing platform is usually required to the requirements for info exploration and analytic actions in IoT. In this unique newspaper, writer suggested such a system established on Joint Cloud Stop string and research a new circumstance of wise journeying based mostly on the recommended system [16].

Wearable skilled products due to sensor constantly bring about tremendous data which is normally known as big info combined with organized and so unstructured data. Anticipated to the difficulty of the info, it is usually hard to course of action as well as , examined the big data pertaining to obtaining useful data that can get beneficial in decision planning. The recommended architecture is utilized for acquiring incorporation of haze computer with cloud work. This architecture as well uses important administration support and info categorization conduct (Sensitive, Crucial and Regular) for offering protection offerings. The platform also utilizes MapReduce founded prediction style to forecast the centre illnesses [17].

Overall performance analysis guidelines many of these as throughput, level of sensitivity, precision, and f-measure will be determined to show the productivity of the offered structures mainly because very well as the conjecture version. Strategies in the agribusiness need to be resolved perishable goods, unstable source variants and strict food security and then durability requisites [18]. The Internet of Things (IoT) could add considerably to resolve these kinds of difficulties as it given permission to be distant managing the area as well as, circumstances of deliveries and so products. The structure backed the supply of inexpensive tailor-made alternatives by way of making use of concept enablers of the Western Long term Net program and by assisting the using again of domain-specific benefits.

III. MQTT PROTOCOL

MQTT is usually among the many virtually all generally utilized methodologies regarding IoT. MQTT allows source confined IoT products to send out, or maybe distribute, details about a provided matter to a machine which usually features seeing that MQTT concept broker. The several brokers after that force the info out to these customers which have recently activated to the theme. To a human being, a subject appears just like a hierarchical document route. Customers may register with a particular position of a topic's structure or perhaps make use of a wild-card personality to sign up to with diverse amounts. All paragraphs must be indented.All paragraphs must be justified, i.e. both left-justified and right-justified.

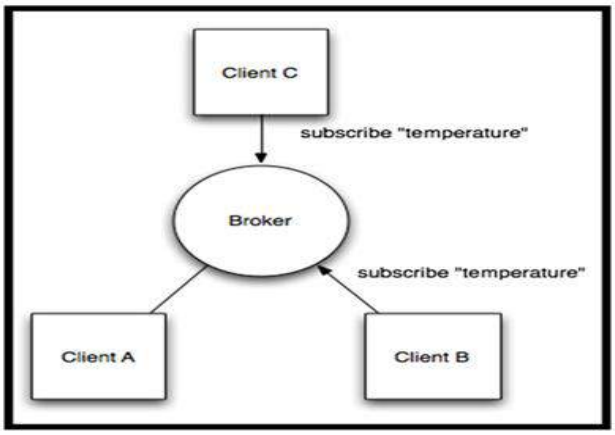


FIGURE 1: MQTT PROTOCOL COMMUNICATION (SOURCE: ECLIPSE)

MQTT can be communication focused. Every message is definitely an under the radar amount of info, opaque with the broker. Each and every message can be posted to addresses, regarded as a matter. Customers might sign up to diverse matters. Every single consumer activated to a theme gets nearly every meaning printed to the issue. For case in point, think about a basic network system with three consumers as well as, a central broker. Almost all clientele will be available with TCP contacts with the broker. Clients B and C subscribe to the subject matter heat. Actually despite the fact that MQTT is certainly engineered to become light-weight, it offers two disadvantages for extremely limited products.

Every MQTT client needs to support TCP and will commonly keep an interconnection wide open to the broker at all occasions. Meant for some conditions just where box loss is usually large or perhaps processing resources will be hard to find, this is normally an issue. MQTT topic names will be frequently lengthy strings that help to make all of them improper for 802.15.4. Together involved with these kinds of shortcomings will be resolved by means of the MQTT-SN process, which identifies an UDP mapping of MQTT as well as , provides broker assist meant for indexing subject titles.

IV. CONCLUSION

Considering that the MQTT protocol turned out to be not really engineered due to protection in brain, the process offers traditionally were utilized in protected back-end systems pertaining to application-specific reasons. MQTT's matter framework can simply type a large tree, and so there's no obvious method to separate a tree into smaller sized reasonable domain names which usually may end up being federated. This creates it hard to produce an internationally scalable MQTT network system since, seeing that the specifications among the theme tree develop, the difficulty raises. , As a future research, Top quality concerning Solutions requires to get recognized. QoS relates to the contract between the sender of a concept as well as, the message's receiver. QoS will determine the assurance of performance in mentioning to a particular concept. QoS functions as an important characteristic in MQTT, providing the customer the capability to select amongst 3 amounts of support.

V. REFERENCES

[1]. Kawaguchi, Ryo, and Masaki Bandai. "Edge Based MQTT Broker Architecture for Geographical IoT Applications." 2020 International Conference on Information Networking (ICOIN). IEEE, 2020.

[2]. Aghenta, Lawrence O., and M. Tariq Iqbal. "Design and implementation of a low-cost, open source IoT-based SCADA system using ESP32 with OLED, ThingsBoard and MQTT protocol." AIMS Electronics and Electrical Engineering 4.1 (2020): 57.

[3]. Aishwarya, K., and R. Manjesh. "A Novel Technique for Vehicle Theft Detection System Using MQTT on IoT." International Conference on Communication, Computing and Electronics Systems. Springer, Singapore, 2020.

[4]. Terada, Keitaro, et al. "Proposal of MQTT distributed broker control mechanism." 2020

International Conference on Information Networking (ICOIN). IEEE, 2020.

[5]. Mukherjee, Amartya, Nilanjan Dey, and Debashis De. "EdgeDrone: QoS aware MQTT middleware for mobile edge computing in opportunistic Internet of Drone Things." Computer Communications 152 (2020): 93-108.

[6]. Chien, Hung-Yu, et al. "A MQTT-API-compatible IoT security-enhanced platform." International Journal of Sensor Networks 32.1 (2020): 54-68.

[7]. Markovic, Milan, and Peter Edwards. "Enhancing Transparency of MQTT Brokers For IoT Applications Through Provenance Streams." Proceedings of the 6th International Workshop on Middleware and Applications for the Internet of Things. 2019.

[8]. Lo, Nai-Wei, and Sheng-Hsiang Hsu. "A Secure IoT Firmware Update Framework Based on MQTT Protocol." International Conference on Information Systems Architecture and Technology. Springer, Cham, 2019.

[9]. Rodríguez, Alejandro, Lars Michael Kristensen, and Adrian Rutle. "Formal Modelling and Incremental Verification of the MQTT IoT Protocol." Transactions on Petri Nets and Other Models of Concurrency XIV. Springer, Berlin, Heidelberg, 2019. 126-145.

[10]. Kashyap, Monika, Vidushi Sharma, and Neeti Gupta. "Taking MQTT and NodeMcu to IOT: Communication in Internet of Things." Procedia computer science 132 (2018): 1611-1618.

[11]. Oliveira, Guilherme MB, et al. "Comparison between MQTT and WebSocket protocols for Iot applications using ESP8266." 2018 Workshop on Metrology for Industry 4.0 and IoT. IEEE, 2018.

[12]. Roy, Deepsubhra Guha, et al. "Application-aware end-to-end delay and message loss estimation in Internet of Things (IoT)—MQTT-SN protocols." Future Generation Computer Systems 89 (2018): 300-316.

[13]. Ansari, Danish Bilal, A. Rehman, and R. Ali. "Internet of Things (IoT) Protocols: A Brief Exploration of MQTT and CoAP." International Journal of Computer Applications 975 (2018): 8887.

[14]. Kang, Do-Hun, et al. "Room temperature control and fire alarm/suppression IoT service using MQTT on AWS." 2017 International Conference on Platform Technology and Service (PlatCon). IEEE, 2017.

[15]. Firdous, Syed Naeem, et al. "Modelling and evaluation of malicious attacks against the iot mqtt protocol." 2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData). IEEE, 2017.

[16]. Kim, Sung-Jin, et al. "A study on adaptive QoS control system based on MQTT for reducing network traffic." International Conference on Future Information & Communication Engineering. Vol. 9. No. 1. 2017.

[17]. Handosa, Mohamed, Denis Gračanin, and Hicham G. Elmongui. "Performance evaluation of MQTT-based internet of things systems." 2017 Winter Simulation Conference (WSC). IEEE, 2017.

[18]. Kim, Sung-jin, and Chang-heon Oh. "Method for Message Processing According to Priority in MQTT Broker." Journal of the Korea Institute of Information and Communication Engineering 21.7 (2017): 1320-1326.Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[19]. M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Theoretical Study of Analysis Prevention and Detection of different Financial Services of Data Mining

Geetanjali Rohan Kalme[1], Dr. Manoj S. Kathane[2]
*[1]IT, Mumbai University, A. P. Shah Institute of Technology, Thane, Maharashtra, India
[2]Electronics and Telecommunication, TSEC, Burhanpur, Madhya Pradesh, India

ABSTRACT

The Cyberspace in India is developing swiftly. It has presented ascend to new open entrances in each pitch we can consider - be it change, business, sports or training. There are dissimilar borders to a money. Network likewise has its own weaknesses. One of the significant disservices is Cybercrime illegal behavior carried out on the web. The network, alongside its drawbacks, has likewise presented us to security hazards that accompany associating with an enormous system. PCs are being abused for illegal operations like email secret work, Visa misrepresentation, suitable, program design theft, mental pestering, etc., which incidence our protection and scandal our abilities. Crimes in the internet are on the mounting. Building up a economic digital wrongdoing recognition framework is a difficult assignment and secure public activity. At whatever point any online exchange is performed through the charge card, at that point there isn't any framework that without a doubt predicts an exchange as fake. It just predicts the probability of the exchange to be a false.

Keywords: Data Mining Techniques, Data Mining Tasks, Data Mining Applications, Clustering, Classification.

I. INTRODUCTION

Evidence digging is utilized for analyzing basic evidence, including agreements numbers, budgets, and clients, to grow better-promoting systems, improve the presentation or diminishing the expenditures of maintaining the business. Similarly, Data mining serves to find new instances of behaviour among customs [2]. It includes collection, mining, investigation, and statistics of information. Data Mining may additionally be explained as a logical method of finding helpful info to seek out helpful knowledge. Once you discover the knowledge and patterns, data processing is employed for creating choices for developing the business.

The most important steps concerned within the process are:

(i) Extract, modify and consignment knowledge into a knowledge warehouse.

(ii) Store and manage knowledge in an exceptionally third-dimensional info.

(iii) proposal knowledge access to business analyst's mistreatment application mainframe code.

II. USES OF DATA MINING

Data mining is used for inspecting unprepared data, which includes auctions numbers, prices, and customers, to rise better publicity and promotion

strategies, improve the overall performance or minimize the costs of jogging the business. Also, Data mining attends to discover new patterns of conduct amongst consumers [3].

Data Mining is used for predictive and descriptive analysis in business:

(i) The resulting sample in Data Mining is supportive in a better hold of client behaviour, which leads to advanced & productive future results.

(ii) Data Mining is used for finding the secreted records through coming near the market, which is advisable for the enterprise however has not but reached.

(iii) It is similarly used for recognizing the region of the market, to gain advertising dreams and generate a moderately exact ROI.

(iv) Data Mining helps in transporting down functioning price, through discovering and defining the practicable areas of investment.

A. Benefits or Advantages of Data Mining Techniques:

There are a few categories of advantages and points of notice of information mining frameworks.

One of the basic issues of these mining makes a total structure of examination of mining methods [4].

1. It is helpful to guess upcoming trends:

Utmost of the working nature of the data mining systems carry on all the informational factors of the elements and their structure.

One of the common benefits that can be derived from these data mining systems is that they can be helpful while predicting future trends. And that is quite possible with the help of technology and behavioral changes adopted by the people.

2. It shows customer behaviors:

There are a few categories of advantages and points of awareness of data mining frameworks. One of the basic problems of these mining makes a total structure of examination of mining approaches.

3. Supports in decision making:

There are some human beings who make use of these data mining strategies to assist them with some type of decision making. At the present time, all the chronicles about approximately can be decided without any difficulty with the involvement of technology and similarly, with the support of such skill one can make a unique decision about somewhat unidentified and unpredicted.

4. Increase company revenue:

As it has been well-defined in the earlier that data mining is a system whereby which it includes some kind of technology to gather some records about anything possible.

This kind of discipline makes matters simpler for their income proportion. As people can acquire facts about the promoted merchandise online, which finally reduces the value of the product and their services.

B. Limitations or Disadvantages of Data Mining Techniques:

Data mining technology is approximately that helps one person in their decision making and that verdict making is a process wherein which all the factors of mining is complicated exactly [4]. And while the involvement of these mining systems, one can come across several disadvantages of data mining and they are as follows.

1) It interrupts user secrecy:

It is a known fact that data mining gathers information about people using some market-based techniques and information technology. And these data mining process involves several numbers of factors.

But while including those aspects, data mining system interrupts the privacy of its user and that is why it lacks in the matters of safety and security of its users. Ultimately, it creates miscommunication between people.

2) Additional unsuitable data:

The key functions of the data mining systems generate a relevant space for valuable information. But the main problem with these information varieties is that there is a possibility that the collection of information processes can be a slight overwhelming for all.

Consequently, it is very much crucial to maintain a minimum level of limit for all the data mining techniques.

3) Misuse of evidence:

As it has been clarified earlier that in the data mining system the opportunity of safety and security measure are actually nominal. And that is why some can misuse this information to harm others in their own way.

Consequently, the data mining system needs to change its development of working so that it can reduce the ratio of misuse of information through the mining process.

4) Accuracy of data:

Most of the time while collecting information about confident fundamentals one used to seek help from their clients, but currently everything has changed. And now the process of information collection made things easy with the mining technology and their methods. One of the most possible boundaries of this data mining system is that it can provide accurateness of data with its own limits.

III. DATA MINING TECHNIQUES

Basic types of Data Mining techniques are as follows [1]:-

- Predictive
- Descriptive

III.I TYPES OF PREDICTIVE

A. **Classification**
B. **Regression**
C. **Time Series analysis**
D. **Prediction**

A. Classification:

Grouping is a data mining (AI) strategy used to anticipate bundle support for data models. For example, you may wish to use gathering to foresee whether the atmosphere on a particular day will be bright, blustery or cloudy. Standard request methodology consolidates decision trees and neural frameworks.

B. Regression:

Regression is an information mining dimensions that forecasts a number. Benefit, deals, contract rates, house estimations, area, temperature, or separation could all be anticipated utilizing degeneration policies. For instance, a relapse model could be utilized to predict the estimation of a house dependent on the

spot, number of places, part size, and different apparatuses [3]. Regression is an information mining strategy used to fit a condition to a dataset. The least difficult type of relapse, direct relapse, utilizes the equation of a straight line (y = mx + b) and decides the suitable qualities for m and b to foresee the estimation of y dependent on a given estimation of x. Boosted procedures, for example, various relapse, permit the utilization of more than one information variable and take into account the fitting of more intricate models, a quadratic condition.

C. Time Run Analysis:

A period arrangement speaks to an mixture of qualities got from successive assessments after some time. Time-arrangement information mining comes from the craving to rectify our distinguishing capacity to visualize the state of information. People depend on complex plans so as to perform such spending. We can actually abstain from focused on little uncertainties so as to determine a thought of shape and recognize immediately similarities between designs on different time scales.

D. Prediction:

Predicting the credentials of one aspect primarily based purely on the description of another, associated element.

1. Not always future events, simply unknowns
2. Based on the relationship between a portion that you can recognize and an element you want to predict Predictor => Predicted.
3. When constructing a predictive model, you have statistics overlapping each.
4. When the procedure of one, you have facts describing the predictor and you choose it to inform you the expected value.

3.2 TYPES OF DESCRIPTIVE

A. **Clustering**
B. **Summarization**
C. **Association Rules**
D. **Sequence Discovery**

A. Clustering:

Clustering is operated to store the information into clusters as per their qualities, attributes, resemblances and dissimilates. In this methodology same sort information are store in same gatherings and these gatherings are known as groups however information is heterogeneous between two groups. Bunches can be applied on gathering of certain schools to examine the similarities and contrasts between these schools, understudies can be grouped together to anticipate the distinctions in their behavior.

B. Summarization:

Summarization is a key information mining idea which includes procedures for finding a minimized depiction of a dataset. Straightforward synopsis strategies, for example, classifying the mean and standard deviations are frequently applied for information examination, information perception and mechanized report age.

C. Association Rules:

Association rule getting to identify is a famous and properly researched approach for discovering attractive family members between variables in immense databases. It is supposed to discover robust guidelines located in databases the use of exclusive measures of interestingness. Based on the thinking of robust rules.

D. Sequence Discovery:

Sequential Pattern mining is a theme of records mining involved with determining statistically applicable patterns between chronicles examples the place the values are delivered in a sequence. It is generally supposed that the values are discrete, and as

a consequence time sequence mining is intently related, however mostly viewed a distinctive activity. Sequential sample mining is a one of a kind case of structured facts mining.

IV. DATA MINING EXPOSURE METHODS

A. violence as a communal unhealthiness

General violence could be a serious social downside universal. it's expected that third of girls worldwide have knowledgeable some style of violence by their intimate partner in their period (WHO 2017). The National Intimate Partner and Sexual Violence Survey (2011) found that ~35.6% of girls report a period rate of intimate partner abuse of some style of violence, like rape, physical violence, or irritation. even if lady's area unit a lot of possibly to be victims of violence, men are victimized by friendly partners. Nearly 28.5% of men report being the victims of some style of violence by assistant intimate partner in their period. Same-sex intimate partner violence is additionally a heavy public health issue. a 3rd of lesbian ladies (33.5%) and one in four gay men (26%) expertise a minimum of one form of violence in their period. violence is related to hurtful penalties for physical health (e.g., injury, chronic pain), mental state e.g., unhappiness, posttraumatic stress disorder, sexual health (e.g., sexually transmitted diseases), and women's procreative health [5].

B. Information mining application to medicinal services misrepresentation discovery

The medicinal services part is a fascinating objective for fraudsters. The availability of a lot of evidence makes it believable to handle this issue with the adoption of information mining strategies, making the examining procedure more proficient and powerful. This exploration has the target of building up a innovative information mining model dedicated to extortion identification among clinics utilizing Hospital Discharge Charts (HDC) in Administrative Databases. Precisely, it is centeredaround the DRG upcoding practice, i.e., the inclination of enrolling codes for offered types of assistance and inpatients wellbeing status so to make the hospitalization fall inside a more profitable DRG class [6].

C. Data mining for security applications

Data mining has many applications in safekeeping together with in nationwide safety (e.g., investigation) as well as in cyber protection (e.g., virus discovery). The threats to countrywide protection consist of attacking structures and terminating necessary setups such as power grids and cable systems. Data mining techniques are being used to identify doubtful humans and groups, and to discover which folks and groups are capable of carrying out terrorist actions. Cyber security is concerned with protecting machine and community systems from dishonesty due to malicious software together with Trojan horses and viruses. Data mining is also being utilized to provide options such as intrusion detection and auditing [7].

D. Data Mining Application for Cyber Credit-Card Fraud Recognition System

Data mining has popularly earned attention in opposing cyber credit-card fraud due to the fact of its positive artificial Intellect (AI) methods and algorithms that can be applied to notice or predict fraud thru Knowledge Discovery from unfamiliar patterns derived from gathered data [8].

E. Scam Detection System in Mobile Payment Service Using Data Mining

As rising of clever telephone dissemination over the world, various cell payment offerings have been developed and scam communications have significantly improved. Even though many financial agencies have arranged safety options to notice fiddle transactions in on/off-line situation, cellular fee offerings nevertheless lack fraud detection options and researches. The cell charge is mainly covered of micro-payments and fee surroundings is exceptional

from other payments, so mobile-specialized fraud detection is needed [9].

V. TASKS IN SCAM RECOGNITION

Building a scam exposure tool is no longer as easy as it looks. The expert wishes to determine which studying strategy to use (e.g., supervised getting to know or unsupervised learning), which algorithms to use (e.g., Logistic regression, selection trees, etc.), which points to use, and most significantly, how to deal with the class inequality trouble (fake instances are particularly sparse as compared to the legit cases). Class imbalance is now not only the most important challenge in fraud detection system. Overlapping of the authentic and fraudulent instructions due to partial information about the transaction files is some other disturbance in the classification responsibility, and most desktop gaining knowledge of algorithms disappoint underneath these circumstances.

In a real-life situation, a fraud detection dummy predicts the nature of class (genuine or fraudulent) and offers the alert for the most suspicious transaction to the investigators. Detectives then perform a supplementary investigation and grant feedback to the fraud recognition device to enhance its performance. However, this procedure can be an overhead for the investigators due to which uniquely a few transactions are validated on time by the investigators. In such a case, just a few feedbacks are provided to the predictive model, which normally outcomes in a less important accurate dummy.

As regulatory establishments very rarely expose the customer information to the public due to confidentiality issues, the actual economic datasets are very hard to find. This is one of the most important tasks in fraud exposure lookup work

VI. FRAUD RECOGNITION PROCEDURE

The transactions are first checked at the terminal factor to be valid or not, which is proven in figure.1

At the mortal opinion, certain essential conditions such as satisfactory balance, authentic PIN (Personal Identification Number), etc. are authorized and the transactions are filtered accordingly. All the valid transactions are then scored with the aid of the predictive model, which then categorizes the transactions as proper or fraudulent.

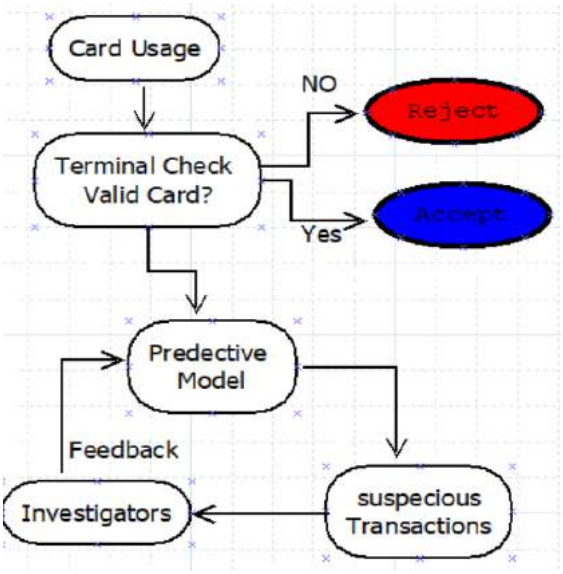


Fig. 1. Fraud recognition procedure

The agents investigate every fake aware and deliver comments to the predictive model to enhance the model's performance[10].

VII. CONCLUSION

We considered dissimilar scam recognition data mining techniques giving to different zones. Data mining is a familiar sector of investigating, expecting and significant guidelines from the huge quantity of information and discovery accurate, earlier unidentified outlines. This paper concentrations on data mining techniques as remarkable method for scam outlines exposure in every zone. Finally, the

bottom line is that all the techniques, techniques and facts mining systems assist in the discovery of new creative things. And at the stop of this discussion about the records mining methodology, one can actually understand the feature, elements, purpose, characteristics, and benefits with its very own limitations. Therefore, after reading all the above-mentioned records about the facts mining techniques, one can decide its credibility and feasibility even better.

VIII. ACKNOWLEDGMENT

The authors would like to thank the editor, the guest editors and all reviewers for valuable comments on an earlier version of the manuscript.

IX. REFERENCES

[1]. Mr. Nilesh kumar dokania & ms. Navneet kaur," comparative study of various techniques in data mining", international journal of engineering sciences & research technology(IJESRT),ISSN:2277-9655, CODEN: IJESS7, IC™ Value: 3.00J.

[2]. https://m.economictimes.com/definition/data-mining/amp

[3]. https://www.digitalvidya.com/blog/what-is-data-mining/

[4]. https://content.wisestep.com/data-mining-purpose-characteristics-benefits-limitations/

[5]. Jia Xue, PhD, Junxiang Chen, PhD, and Richard Gelles, PhD," Using Data Mining Techniques to Examine Domestic Violence Topics on Twitter", VIOLENCE AND GENDER Volume 6, Number 2, 2019[a], MaryAnnLiebert, Inc. DOI:10.1089/vio.2017.

[6]. Michela Carlotta Massi, Francesca Ieva and Emanuele Lettieri, "Data mining application to healthcare fraud detection: a two-step unsupervised.

[7]. clustering method for outlier detection with administrative databases", BMC Medical Informatics and DecisionMaking (2020) 20:160 https://doi.org/10.1186/s12911-020-01143-9

[8]. U.U.Veerendra, B.Ravitheja and K.Veeresh, "DATA MINING FOR SECURITY APPLICATIONS", International Journal Of Core Engineering & Management (ISSN: 2348-9510) Special Issue, NCETICT -2017

[9]. John Akhilomen," Data Mining Application for Cyber Credit-Card Fraud Detection System", P. Perner (Ed.): ICDM 2013, LNAI 7987, pp. 218–228, 2013.© Springer-Verlag Berlin Heidelberg 2013

[10]. Hee Chan Ha,n Hana Kim and Huy Kang Kim," Fraud Detection System in Mobile Payment Service Using Data Mining", Journal of The Korea Institute of Information Security & Cryptology VOL.26, NO.6, Dec. 2016

[11]. Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

TECHNO SCIENCE ACADEMY
THE INTERNATIONAL OPEN ACCESS PUBLISHER

D Y PATIL GROUP
Dr D Y Patil School of Engineering, Lohegaon, Pune

IJSRST

FIRST INTERNATIONAL CONFERENCE ON COMPUTER ENGINEERING DYP-ICCE 2020

Organised by

Dr. D. Y. Patil School of Engineering
Dr Ajeenkya DY Patil Knowledge City Lohegaon, Pune, Maharashtra, India
Department of Computer Engineering and Association of Computer Engineering in Association with
International Journal of Scientific Research in Science and Technology
Sponsored By
TechnoScience Academy (The International open Access Publisher)

Publisher

Technoscience Academy

Website : www.technoscienceacademy.com
Email: info@technoscienceacademy.com