Online ISSN : 2395-602X
Print ISSN : 2395-6011
www.ijsrst.com

DOI : https://doi.org/10.32628/IJSRST.SPIN-2K24

Conference Proceedings

1st International Conference on Security, Parallel Processing, Image Processing and Networking [SPIN-2K24]

26th March 2024

Organized By

Department of Computer Science & Engineering,
Viswajyothi College of Engineering and Technology,
Muvattupuzha, Ernakulam, Kerala, India
In Association With
CSI, ISTE, R&D & A2Z Edulearning Hub

VOLUME 11, ISSUE 14, MARCH-APRIL-2024

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN SCIENCE AND TECHNOLOGY

PEER REVIEWED AND REFEREED INTERNATIONAL SCIENTIFIC RESEARCH JOURNAL

Scientific Journal Impact Factor : 8.014

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

1st International Conference on Security, Parallel Processing, Image Processing and Networking [SPIN-2K24]

26th March 2024

Organized by

Department of Computer Science & Engineering, Viswajyothi College of Engineering and Technology, Vazhakulam, Muvattupuzha, Ernakulam, Kerala, India

In Association With

CSI, ISTE, R&D & A2Z Edulearning Hub

Published By

International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 Online ISSN : 2395-602X

Volume 11, Issue 14, March-April-2024

International Peer Reviewed, Open Access Journal

Published By
Technoscience Academy

website: www.technoscienceacademy.com

CONTENT

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

Advancements and Comparative Analysis of Text-to-SQL Techniques: A Comprehensive Review

Mariyam NP, Asst. Prof. Savyan PV
*Department of Computer Science, Government Engineering College, Palakkad, Kerala, India

ABSTRACT

The increasing need for natural language interfaces to databases (NLIDBs) has led to a notable surge in interest in text-to-SQL technology in recent years. The growing number of applications that require seamless integration between natural language queries and database operations highlights the need for robust text-to-SQL solutions that can bridge the gap between human language and database systems. In order to evaluate the benefits, drawbacks, and efficacy of several text-to-SQL technologies, this paper provides a comprehensive comparison analysis of those systems using a variety of datasets and assessment criteria. Through the integration of data from several methodologies and empirical assessments, this study aims to drive future research directions in this rapidly evolving field and offer valuable insights into the state-of-the-art in text-to-SQL technology

Index Terms:—Natural Language Processing, Structured Query language, Text to sql task, LLM

I. INTRODUCTION

In today's data-driven economy, seamless interaction between users and databases is critical foreffective information extraction and decision-making. The need for user-friendly interfaces thatenable users to query databases in plain English without the need for Structured Query English(SQL) proficiencystemsfrom thegrowing volume andcomplexityof data. This demandhasgiven rise to text-to-SQL technology, a name for Natural Language Processing (NLP) techniques.These methods are particularly made to convert text that can be read by humans into SQLqueries.Natural Language Processing (NLP) is a subfield of artificial intelligence that aims toenable computers to understand, translate, and generate human language. It handles severaldifferent jobs, including as sentiment analysis, language translation, and text summarization.Intext-to-SQL, natural language queries are evaluated and interpreted using NLP algorithms, whichthen extract the underlying purpose and convert it into SQL instructions that database systemscan understand.Structured Query Language, or SQL, is the industry standard language for man-aging and interacting with relational databases. It provides a comprehensive command set formanaging, querying, and modifying structured data. SQL queries frequently include specificationsfor operations like data selection, filtering, and aggregation in addition to creating relationshipsbetween different database fields.The aim of text-to-SQL is the automatic translation of

naturallanguage queries into equivalent SQL queries. This process requires an understanding of theinputtext'ssyntacticandsemanticstructures,domain-specificterminology,anddatabaseformat.

II. OVERVIEW OF EXISTING TEXT-TO-SQL APPROACHES

Thepurposeoftext-to-SQL,anaturallanguageprocessing(NLP)operation,istotranslatenaturallanguage inquiries or questions into Structured Query Language (SQL) queries that may be runagainst a database to obtain pertinent data. In order to create systems that can comprehend andreact to user queries in a database-driven application, this work is crucial. An outline of current methods for text-to-SQLoperationsisprovidedbelow: MethodsBasedonRules[1][2]:Conventional rule-based approaches use lexical patterns and thoughtfully crafted grammar rules to convertnatural language queries into SQL queries. Since these systems often need a high level of topicexpertise and human effort for rule establishment and updating, they are less scalable and flexi-ble.Methods Based on Templates[6]: By entering data from the natural language inquiry into pre-made templates, these approaches generate SQL queries. Templates may contain placeholdersfor conditions, database table names, column names, and other data. Although template-basedmethods are more flexible than rule-based methods, they may not be able to handle variations innaturallanguageinputandstillrequirethecreationofmanualtemplates.Supervisedlearningmeth-ods utilize make use of annotated datasets in which every inquiry in natural language is matchedwithaSQLquery.NeuralNetworkBasedTechniques[3][4][5]Deeplearningapproaches,Templatebasedappr oaches,Hybridapproaches,LLMbasedapproachesarewidelyusedfortexttosqltask.

III. COMPARATIVE STUDY OF TEXT-TO-SQL TECHNIQUES

3.1 Rule Based Approach for text to sql task

The paper[1] presents an effective technique for automating the conversion of Natural LanguageQueries into SQL Queries.The proposed model uses Natural Language Processing (NLP) tech-niques to achieve this purpose. The technique is designed specifically for TP officials who workwith student databases but lack SQL knowledge. It offers a structure for converting Englishquestions from TP officers into SQL queries. Users can talk into the system, and it will translatetheir questions into text format before converting them into SQL queries. This feature enhancesaccessibilityandusability.Theproposedsystemarchitectureconsistsoffourmainsteps:Thepro-cess of breaking up the query into distinct words is called tokenization.The technique of mappingtokens to database terms using a dictionary is called lexical analysis.Tokens are connected todictionaryattributesthroughtheprocessofsyntacticanalysis,whichenablesSemanticAnalysistoprocessthem.Sema nticanalysisistheprocessofidentifyingconditionsorsymbolsinthequeryand mapping them to the lexicon.The paper's conclusion highlights that the proposed techniquemay facilitate database interaction for users—particularly who are not familiar with SQL. It explainsthe steps that are needed in answering questions and emphasizes how important it is to addressproblems like ambiguity in natural language queries. A potential shortcoming of the proposedapproach [1] is its limited vocabulary coverage.The effectiveness of the model primarily dependson the completeness of the dictionary used to convert natural language tokens into SQL querycomponents. If the dictionary lacks synonyms for specific terms or does not have a broad enoughvocabulary,themodelmayhaveproblemsaccuratelyreadingandtranslatingrequests.Thiscouldlead to errors

or inadequate outcomes. Paper[2][2019] uses Natural Language Processing (NLP)to convert structurednaturallanguagequestions into SQLqueries, facilitatingsimpler accessto information from a train reservation database. Tokenization, lemmatization, parsing, mapping,and parts of speech tagging are just a few of the NLP techniques that are employed in theprocess. The algorithm achieves a high accuracy rate of 98.89 percentage using a dataset of2880 structured natural language queries on rail ticket and seat availability. Lemmatization, parts-of-speech tagging, regular expression parsing, and tokenization of the input question are the firststeps in the multi-step process used by the proposed system [2]. Semantic analysis is then usedto infer the context of the tokens, which is particularly helpful for figuring out where informationoriginates and ends up when searching for train tickets.The final stage that pulls relevant datarequiredtobuildtheSQLqueryisattributeidentification.TheNLPandmappingphasesmakeupthesystemmodel.BeforeaSQLqueryiscreated,incomingdatamustundergotokenization, lemmatization,parsing,andPOStagging.TheSQLquerygeneratingalgorithmconsistsofseveralstages, including attribute identification, POS tagging, tokenization, lemmatization, and parsing.The system's inability to handle complex database structures or a wide range of query types isone of its drawbacks. It may struggle with complex joins, nested clauses, and several conditionqueries,producingSQLqueriesthatareeithererroneousorincomplete.

3.2 DeeplearningApproachforTexttosqltask

Neural network recurrent architectures Previously believed to be the most effective method fornaturallanguageprocessing(NLP),recurrentneuralnetworks(RNNs)havelatelybeenovertakenby potent transformers. The primary benefit of RNNs is its capacity to: (a) efficiently handlesequenceinputs,suchawordseriesknownasanNLQ;and(b)generateasequenceoutput,liketheconditionvalueofaWHEREclauseoragroupofgrammarrulesthatcouldleadtoaSQL query. The GRU (Gated Recurrent Unit) and LSTM (Long Short-Term Memory) are two well-liked RNNdesigns.For inputencoding, early systemslike Seq2SQL[7]andSQLNet[8]alsousedLSTMsinadditiontopre-trainedwordembeddings.PretrainedLanguageModelsarecurrently moreeffective for this kind of application. RNNsare still employed to help LMswithinput encoding and to give non-NL series outputs, even though their use in the input encodingphasehassignificantlydecreasedasaresultoftherecentsuccessofTransformersandLanguageModels. To build single-token representations for columns and tables with several words in theirnames(andmultipletokenstorepresentthem),IRNet[9]employsBERTtoencodetheinputNLQand schema. GuideSQL is a novel method for addressing the task of Text-toSQL, which involvesgenerating SQL queries from natural language utterances. The approach, called GuideSQL, isdesignedtoimprovetheaccuracyofpredictingSQLqueriesbyleveraginginformationfromtablesand employing a pruning algorithm to mitigate errors caused by misprediction of table-columndependencies. Predicting tables first is a crucial part of the GuideSQL[3] technique. GuideSQLbases its predictions about tables on the input utterance before directly predicting columns. Thishelps to reduce the prediction space for columns and reduces the likelihood of errors.Once thetables have been predicted, a pruning process is used to remove columns that do not belong inthe predicted tables. By optimizing the prediction process, this step contributes to the provision ofmore accurate SQL queries.A string-matching method is used by GuideSQL to select the mostsensible SQL query out of many predictions. This method can be used to rerank SQL queries thataregeneratedbyanticipatingthetop-Ktables.GuideSQLusesatoincreasetherelevancybetweenutterances and

schemas.To improve the relevance between utterances and schemas, GuideSQLmakes use of a type linking technique. By identifying the kind of entities (TABLE, COLUMN)specified in the query, this method adjusts schema embeddings appropriately.Context-dependentinformation from SQL queries is captured using previous query attention, which is very helpful formanagingintricatedatasetswithseveralcontexts,suchasSParC.TheGuideSQLmethodisbasedontheneuralnetwor kCDSeq2Seqmodelinparticular[3].Thetitle"Context-DependentSeq2Seq,"or CD-Seq2Seq, refers to the way that it considers contextual dependencies when generatingSQL queries from natural language speech. GuideSQL builds on this design by introducing newcomponentsincludingthetypelinkingmethod,previousqueryattention,andguidancemechanismto increase efficiency. These components are integrated into the neural network's architecture toimprove the predictive capacity of SQL queries. The model presented in the paper[5] translatesnatural language questions into SQL queries by leveraging deep learning techniques, such asWord2Vec embeddings and LSTM (Long Short-Term Memory) networks. Furthermore, the modelintegrates fuzzy decision and fuzzy semantic processes to enhance decision-making capabilitiesand manage intricate semantic queries in many areas.Important elements of the strategy consistof:The F-SemtoSql (fuzzy semantic to structured query language) model: This is the primaryneural network model intended to handle the task of generating text from SQL. For training oninput word vectors, it makes use of Word2Vec embeddings and LSTM networks.Dependencygraph approach: By creating semantic dependency graphs, this method helps transform the SQLstatementgenerationproblemintoslotfilling.Attentionmechanism:Usedinconventionalmodelsto solvetheorderproblem,theattentionmechanismimprovesthemodel'scapacitytoconcentrateonpertinent portions of the input when making decisions.Division of complicated tasks: The methodusesarandommaskingtechnique fortrainingandsplitscomplextext-to-SQLjobsintofourlevels:atomic events, aggregate events, complex events, and composite events.Fuzzy choice module:To address uncertainty and variability in the input, this module includes fuzzy decision-makingbasedonconnectionnumbers.

3.3 Neural network based Approach for texttosqltask

In the paper[4], it is discussed how important self-supervised learning is for deep learning, es-pecially forapplicationswhere labeled dataisscarce, such asText-to-SQL. It drawsattentionto the differences between task-specific objectives and self-supervised learning objectives. Theauthors suggest a self-supervised learning system designed specifically for Text-to-SQL jobs inorder to solve this.Their methodology trains models to learn header-column alignment tasks fromunlabeledtabledatabytakinguseoftheintrinsicfeaturesofText-to-SQLjobsandthearchitectureoftablecontents.Then,byusingannotatedexamplesduringsupervisedText-to-SQLtraining,thisinformation is applied to improve the model's prediction of SQL statements. One limitation of thesuggested model[4] is its ability to be applied to new areas. Although the method could work wellon the datasets it was trained on, there's a chance it won't generalize as well to other, untestedareas. When faced with new table structures, query formats, or linguistic variants not found in thetrainingsetofdata,themodelcouldfinditdifficulttoadjust.

3.4 LargeLanguageModelfortexttosqltask

Thepaperinvestigates[11]thefundamentalproblemofdeterminingtheequivalencebetweentwo SQL queries, which is crucial in various data management applications and SQL generationtasks, such as evaluating the quality of

generated SQL queries in textto-SQL tasks. In order toassess thequality ofgenerated SQLqueries in text-to-SQL activities, forexample, andforusein a variety of data management systems, the study [11]looks at the fundamental challenge ofassessing the equivalency between two SQL queries. Because SQL equivalency is so difficult,there are nocomprehensive solutions, even afterdecades of research. Large LanguageMod-els(LLMs)[11]haveadvancedrecently,demonstratingsignificantreasoningabilityinarangeof activitieslikeasquestionanswering, discussion, andproblemsolving.The purposeof thestudy is to investigate whether LLMs can reliably ascertain whether two concepts of SQL queryequivalency—semantic equivalency and relaxed equivalency—are met.The study offers MiniatureMull and Explain Compare as two prompting strategies to help LLMs produce high-quality replies.The Miniature Mull method [11] asks LLMs to run queries on a basic database instance andinvestigate whether a counterexample can be found by making changes to the database in ordertoassesssemanticequivalency.Inordertoassessrelaxedequivalency,theExplainCompare[11]approach asks LLMs to clarify the inquiries and compares them to see whether there are anynotablelogicaldisparities.Thefindingsoftheexperimentsindicatethat LLMshavepromisingskillstodetectsemanti cequivalencyofSQLquerieswhenpresentedwiththesuggestedmethodologies.But difficulties still exist, suggesting that more study in this field is required. According to thestudy, determining SQL equivalency using LLMs may be a more useful criterion for assessingSQL creation than more conventional techniques like execution correctness. The work of usingprompting techniques to convert natural language text into SQL queries—a process called as text-to-SQL—is discussed in this section[10]. Although earlier research on text-to-SQL prompting waslimitedtoevaluatingitinazero-shotenvironment,itshouldberememberedthatforthemajorityof activities, zero shot prompting may only offer a lower constraint on the possible capabilities ofLanguage Models (LLMs). The suggested approach is demonstrated to perform noticeably betterthan few-shot prompting techniques.On the Spider dataset, comparative results are shown be-tween inference-only and fine-tuning techniques, using precise set match accuracy and executionaccuracy as evaluation criteria. The presented table presents a performance comparison on thedevelopmentsetofSpiderbetweenzero-shotandfew-shotpromptingmethodsandoptimized techniques.Thepaperalsoshowsgainsinexecutionaccuracyandprecisesetmatchaccuracyby comparing the proposed method with existing approaches on the cross-domain benchmarks,Spider andBIRD.Thesuggested solutionuses GPT-4 andCodeXDavinci models to achievehigh execution accuracy and accurate set match accuracy on the Spider dataset. The suggestedmethod's few-shot in-context nature is the reason for the difference between precise match andexecution accuracies. In spite of this, the approach outperforms earlier state-of-the-art (SOTA)findings on the Spider dataset, setting new benchmarks in execution accuracy.The suggestedmethodestablishesanewSOTAresultinexecutionaccuracyontheBIRDbenchmark,provingits efficacy with a legitimate efficiency score. The technique demonstrates its effectiveness bysurpassing a GPT-4 baseline. Overall, these results highlight the improvements and potency oftheproposedmodel.

IV. EVALUATION METRICS AND CRITERIA FOR DIFFERENT TEXT-TO-SQLAPPROACHES

4.1 SemanticAccuracy

This metric focuses on how well the generated SQL query captures the intended meaning of thenaturallanguagequestion

4.2 QuestionMatchingAccuracy

This metric would typically assess how well the system matches the natural language question(input)tothecorrespondingSQLquery(output).

4.3 InteractionMatchingAccuracy

Itreferstohowwellthesystemhandlesinteractions.

4.4 ExactMatchAccuracy(EM)

ItcheckswhetherthegeneratedSQLquerymatchesthegoldstandardqueryexactly.

4.5 ExecutionTime

The total amount of time that the system spends prepping, parsing, and actually executing SQLqueries

4.6 Scalability

The model's capacity to manage big datasets and intricate queries without seeing appreciableperformancedeterioration.

4.7 ErrorAnalysis

Error analysis is the process of locating and examining mistakes that the system has made, suchaslogical,syntactic,andsemanticmistakesintheSQLqueriesthatareproduced.

4.8 Generalization

Ameasureofamodel'sresilienceandsuitabilityforreal-worldsituations,itexpresseshowwellitcangeneralizetopreviouslyuntesteddataordomains.

4.9 Adaptability

The system's capacity to adjust without the need for extensive retraining or revisions in responsetochangesintheunderlyingdatabaseschema,querypatterns,ornaturallanguagevariances

4.10 HumanEvaluation

AsubjectiveevaluationbydomainexpertsorendusersoftheusabilityandqualityoftheproducedSQLqueries,offeringimportantinsightsintoreal-worldperformance

Model	Approach	Performance Metrics	Dataset	Performance in percentage
Model proposed in [2]	Rule based Approach	Accuracy	Test.set(internal)	98.89
Self supervised model[4]	Neural network based approach	Exact Match Accuracy	Spider	45.25
F-Semtosql[3]	Deep learning Approach	ACC_n(atomic events, composite events, aggregated events, and complex events) assesment	Spider Dev.set	37.61
IRnet[9]	Deep learning approach	Exact match Accuracy	Spider Test.set	39.5
Guidesql[5]	Deep learning Approach	Question matching accuracy	Spider Dev.set	49.2
Seq2Sql[7]	Deep learning approach	ACC_{ex}	Wikisql Test.set	59.4
Seqnet[8]	Deep learning approach	ACC_{ex}	Wikisql Test.set	70.3
Few-short Gpt-4 model[10]	LLM based model	Execution Accuracy	Spider Dev.set	67.4

Fig.1.Comparisontable,ACCn=(N_Acc^n)/NnTheNnisthetotalnumberofdifferentlevelsofdatain the test set .The N_Acc^n is the number of SQL statements which generated by the model to matchtheactualstring.ACCex=(Nex)/NNistotalnumberofexamplesinthedataset,Nexthenumberofqueriesgivescorrectresult

V. PERFORMANCE EVALUATION AND ANALYSIS

Fortext-to-SQLjobs,largelanguagemodels(LLMs)haveimprovedonpreviousmethodsinanum-ber of ways.Performance criterias include better comprehension of natural language,Decreasedreliance on manually created rules,Handling ambiguity and variability,Scalability and adaptability.LLMs are better able to grasp natural language questions because to their extensive pre-trainedknowledge and excellent language understanding capabilities. They enable more accurate inter-pretations of user inquiries by capturing complex subtleties, context, and variances in linguisticexpressions. Conventionalmethodsforconverting naturallanguage to SQLqueries frequentlyrelyonmanuallycreatedrulesortemplates.Conversely,LLMsacquiretheabilitytoproduceSQL queries straight from natural language input by means of end-to-end training. LearningMachines that have been trained on extensive and varied datasets exhibit good domain and taskgeneralization. Because of this, they can manage a variety of queries and databases withoutrequiring a lot of domain-specific modification or training.It might be difficult for conventionalsystems to correctly comprehend natural language inquiries since they can be ambiguous andvaried. By using their contextual knowledge and probabilistic modeling skills to determine themost likely SQL interpretation given the input, LLMs are excellent at managing ambiguity andunpredictability.LLMs have a great degree of scalability and adaptability, and they can continuallylearn from fresh data and get better.The Table(fig 1) below categorize models along with theirperformanceinvariousdatasetwiththespecificperformancemetrics.

VI. CONCLUSION

Thispaperprovidesathoroughinvestigationintotheevolvinglandscapeoftext-to-SQLtechniques, presenting a detailed analysis of the advancements and methodologies that have shaped thisfield. Through meticulous examination, we have highlighted the substantial progress facilitated bythe integration of cutting-edge technologies such as Large Language Models (LLMs) and novelpromptingstrategies.Theseinnovationshavenotonlybolsterednaturallanguagecomprehensionbut have also substantially reduced the reliance on labor-intensive manual rule creation, fosteringamoreadaptableandscalableframeworkfortext-to-SQLsystems.

Despite these remarkable advancements, persistent challenges confront LLMs when taskedwith generating SQL queries from natural language inputs. The intricate and varied structure ofSQLqueriesoftenposesdifficultiesforLLMs,astheynavigatetheintricatesyntaxrulesinherentindatabase query languages. Striking a delicate balance between syntactic precision and semanticfidelity remains a formidable obstacle in the quest for accurate translation. Bridging the semanticdividebetweenthenuancedexpressionsofnaturallanguagequeriesandthestructuredformatof SQL queries is essential, yet often proves challenging for LLMs, leading to potential inaccuraciesinquerygeneration.

Inparticular,thecomplexitiesinherentincomprehendingthesubtlenuancesofcomplexqueriesor phrases present significant hurdles for LLMs. These challenges underscore the ongoing needfor continued research and innovation to refine LLM capabilities in the text-to-SQL domain. Byaddressing these obstacles head-on, researchers can further enhance the effectiveness andreliability of text-to-SQL systems, paving the way for more seamless interaction between usersanddatabases.

VII. REFERENCES

[1] Conversion of Natural Language Query to SQL Query Abhilasha Kate Satish Kamble Aishwa Published in: 2018SecondInternationalConferenceonElectronics,CommunicationandAerospaceTechnology(ICECA),publisher:IEEEDateAddedtoIEEEXplore:30September2018

[2] Formationof SQL from NaturalLanguageQueryusingNLP UmaM Sneha V SnehaGBhuvana JBharathiComputer Science andEngineering SSN College of Engineering Chennai,IndiaSecond InternationalConferenceonComputationalIntelligenceinDataScience(ICCIDS-2019)B978-1-5386-9471-8/19/31.002019IEEE

[3] A Comprehensive Exploration on Spider with Fuzzy Decision Text-to-SQL Model Qing Li , Member, IEEE, Lili Li , QiLi,andJiangZhongIEEETRANSACTIONSONINDUSTRIALINFORMATICS,VOL.16,NO.4,APRIL2020

[4] Self-supervisedText-to-SQLLearningWithHeaderAlignmentTrainingDonggyuKim1 SeanieLee2Kookmin Bank1,KAIST2 ,South Korea donggyukimc@gmail.com,lsnfamily02@kaist.ac.kr

[5] GuideSQL: Utilizing Tables to Guide the Prediction of Columns for Text-to-SQL Generation Huajie Wang Lei Chen†,Mei Li Mengnan Chen School of Computer Science and Technology, East China Normal Authorized licensed uselimitedto:FondrenLibraryRiceUniversity.DownloadedonOctober26,2020at08:44:09UTCfromIEEEXplore

[6] Importance of Synthesizing High-quality Data for Text-to-SQL Parsing Yiqun Hu, Yiyun Zhao, Jiarong Jiang, WuweiLan,HenryZhu,AnujChauhan,AlexanderLi,LinPan,JunWang,Chung-WeiHang,ShengZhang,JiangGuo,MarvinDong,JoeLilien,PatrickNg,ZhiguoWang,VittorioCastelli,BingXiangAWSAILabsyiyunzhao@arizona.eduyiqunhu,jiarongj, lanwuwei,henghui, chaanj,hanboli, linpan,juwanga, cwhang, zshe, gujiang,mingwd, lilienj, patricng,zhiguow, vittorca, bxiang @amazon.co. Findings of the Association for Computational Linguistics: ACL 2023, pages1327–1343July9-14,2023©2023AssociationforComputationalLinguistics

[7] Zhong, V., Xiong, C., Socher, R.: Seq2sql: generating structured queries from natural language using reinforcementlearning(2017)arXiv:1709.00103v7[cs.CL]9Nov2017

[8] Xu,X.,Liu,C.,Song,D.:Sqlnet:generatingstructuredqueriesfromnaturallanguagewithoutreinforcementlearning.arXiv:1711.04436v1[cs.CL]13Nov2017

[9] Towards Complex Text-to-SQL in Cross-Domain Database with Intermediate Representation Jiaqi Guo1, ZechengZhan2, Yan Gao3, Yan Xiao3, Jian-Guang Lou3Ting Liu1, Dongmei Zhang3 1Xi'an Jiaotong University, Xi'an, China2Beijing University of Posts and Telecommunications, Beijing, China 3Microsoft Research Asia, Beijing, ChinaarXiv:1905.08205v2[cs.CL]29May2019

[10] DIN-SQL: Decomposed In-Context Learning of Text-to-SQL with Self-Correction Mohammadreza Pourreza, DavoodRafieis:arXiv:2304.11015[cs.CL]NeurIPS2023

[11] LLM-SQL-Solver:CanLLMsDetermineSQLEquivalence?FuhengZhao,LawrenceLim,IshtiyaqueAhmadDivyakantAgrawal,AmrElAbbadiUCSantaBarbarafuhengzhao,lawrenceklim,ishitiyaque,divyagrawal

[12] Amer-Yahia,S.,Koutrika,G.,Braschler,M.,Calvanese,D.,Lanti,D.,Lu¨cke-Tieke,H.,Mosca,A.,deFarias,TMendes, Papadopoulos,D.,Patil,Y.,Rull,G.,Smith,E.,Skoutas,D.,Subramanian,S.,Stockinger,K.:Inode:buildinganend-to-enddataexplorationsysteminpractice.SIGMODRec.50(4),23–29(2022)

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

AI-Driven Video Prompt Analysis

Allen Saji*, Ashik David Roy, Nithin V. James, Reenphy George, Prof. Smitha Jacob
*Department of Computer Science, St. Joseph's College of Engineering and Technology, Palai, Kerala, India

ABSTRACT
The vast amount of video data, particularly in surveillance domains, demands an advanced method for efficient content extraction. Traditional approaches face challenges in terms of time, labor, and accessibility. This paper introduces an AI-driven video prompt analysis system, a solution designed for interaction with video data through natural language prompts. The core engine, powered by Python and Yolo V8, enables video analytics, integrating FastAPI, PostgreSQL, and a Python-based video fetcher for video extraction. The user-friendly interface, made with NextJS, TypeScript, and Tauri, ensures an intuitive user experience. The system ensures continuous monitoring, scalability, real-time notifications, and security.
Keywords: Surveillance, AI-driven, Security, Prompt analysis, Video extraction.

I. INTRODUCTION

The ever-expanding increase of video statistics in surveillance domains has propelled the need for superior methodologies in content extraction. Conventional strategies regularly fall short in addressing the demanding situations posed using the sheer extent of records, the labor intensive nature of the analysis, and the shortage of user friendly interfaces. In response to these demanding situations, our project introduces a groundbreaking AI-driven video analysis system. This system seeks to redefine the interplay paradigm with video records by leveraging natural language for content extraction.

The core engine of our system harnesses the power of Python and Yolo V8 for video analytics. To facilitate smooth communication, we integrate fastAPI, PostgreSQL, and a Python-based video fetcher. This combination not only enhances the efficiency of the core engine but also lays the foundation for a responsive and dynamic system. With a user friendly approach, we craft an interface using NextJS, TypeScript, and Tauri, ensuring accessibility for both technical and non-technical users. The subsequent sections will delve into the methodology, and more, providing a detailed idea of the technical factors, key findings, and the broader implications of our work.

II. METHODOLOGY

Our methodology follows a systematic and comprehensive approach to ensure the successful implementation of the AI-driven video prompt analysis system. The key steps in our methodology includes dataset collection, AI model implementation,, real time object detection, performance optimization, user training, ethical

considerations, continuous monitoring and feedback, scalability for large datasets, real time notifications, and user authentication and authorization.

A. Core Engine

The core engine serves as the backbone of our system, leveraging YoloV8 for video analysis and object detection. It analyzes incoming video feeds in real-time or uploaded videos, detecting actions and objects of interest. Once identified, these actions and objects are recorded in the knowledge base for future reference and inference, enabling efficient data retrieval and analysis

B. APIs Integration

Developed using Python with FastAPI, the API integration module handles user requests and data retrieval from the knowledge base. Acting as the intermediary between the frontend, Large Language Models (LLM), and the knowledge base, the backend APIs facilitate seamless communication and data transfer. This component is responsible for managing user interactions, including chat functionalities, and ensuring efficient data exchange between different system modules.

C. Video Fetcher

The Video Fetcher is built using Python and MoviePy library and it plays an important role in processing user prompts and retrieving specific video frames. Upon receiving prompts from users, it communicates with the Large Language Models (LLM) to TRIM relevant video footage based on the specified criteria. By efficiently accessing video data in real time, the Video Fetcher ensures smooth integration with the system, thereby enhancing the overall user experience.

D. Front End Interface

The Frontend Interface serves as the primary interface for users to interact with the system, it offers a platform for accessing and analyzing video data. Developed using NextJS, TypeScript, and Tauri, this enables users to input prompts, upload videos, and access real time CCTV footage with ease. It interacts up close with the system to transmit user prompts and receive processed video frames. Additionally, it interacts with the API Integration to deliver real time updates and notifications to users, ensuring timely and relevant information.

E. Large Language Models

The Language Learning Model (LLM) is an AI system designed to understand and process human language. It's trained on vast amounts of text data to learn patterns and semantics, enabling it to interpret and respond to user queries. In our project, the LLM is fine-tuned to process prompts related to video analysis tasks, such as extracting timestamps or identifying specific actions in video footage.

III. IMPLEMENTATION

A. System Flow

The system flow initiates with users interacting with the Frontend Interface, inputting natural language prompts, or uploading videos. These prompts are processed by the Core Engine, leveraging the YoloV8 model

for video analysis. The API Integration Module facilitates the smooth exchange of data between the Core Engine and external sources. Simultaneously, the Video Fetcher Module responds to user prompts by retrieving and trimming relevant video frames. The coordinated interaction of these modules ensures an integrated and efficient system workflow.

1. **Interaction Between Modules:**
   - 1.1 Frontend Interface Module interacts with the Core Engine Module: Users input prompts and the Core Engine processes them, providing relevant responses through the Front end Interface.
   - 1.2 Integration Module interacts with the Core Engine Module: Facilitates secure communication, enabling data transfer, authentication, and logs between the Core Engine and external entities.
   - 1.3 Video Fetcher Module interacts with the Core Engine and Frontend Interface Modules: Retrieves specific video segments based on user prompts and smoothly delivers them to the Front End Interface.

2. **Process flow:**

The system flow is designed to efficiently process user prompts, analyze real time video footage, and provide relevant outputs. Here's a step-by-step explanation of the flow:

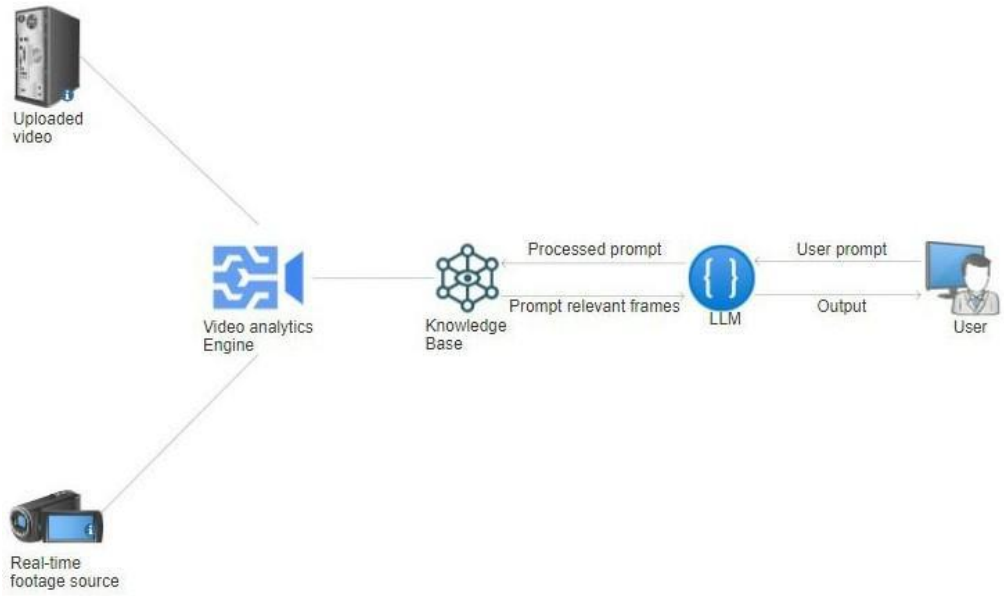


Fig. 1: AI-Driven Video Prompt Analysis System Design

- 2.1. User Input Prompt: The process initiates when a user inputs a prompt, specifying the desired information or action.
- 2.2. Large Language Model (LLM): The user prompt is then processed by the Large Language Model (LLM), which interprets and extracts key information from the natural language input.
- 2.3. Processed Prompt: The processed prompt is generated by the LLM, capturing the details and intent of the user's input.
- 2.4. Knowledge Base: The processed prompt interacts with the Knowledge Base, a storage for information and data relevant to the system.
- 2.5. Video Analytics Engine: Simultaneously, real time footage sources, such as surveillance cameras, are continuously analyzed by the Video Analytics Engine.

2.6. Prompt Relevant Frame Identification: The Video Analytics Engine identifies frames in the real time footage that are relevant to the user's prompt, using video analytics techniques.

2.7. LLM Interaction with Output: The frames identified by the Video Analytics Engine are communicated back to the LLM, improving the understanding of the user prompt.

2.8. User Output: The LLM generates an output based on the processed prompt and relevant frames, presenting meaningful information to the user.

2.9. Loop Completion: The entire process forms a dynamic loop, allowing users to receive outputs based on their prompts and ensuring continuous interaction with real time video analytics.

In summary, these interactions enable a smooth flow of data and commands between different parts of the system, ensuring that users can efficiently input prompts, receive relevant responses, and access the necessary video segments for analysis.

B. Module Implementation

This system's backbone is created by the implementation with a core engine module, which supports video analytics processes. It is a module that has been built using advanced models for object detection and analysis in video frames. Once it receives video data from either CCTV feeds or uploaded files, the core engine initializes object detection models. It identifies objects through careful examination and then sorts them out to be kept alongside timestamps as results. The core engine collaborates with other modules to exchange video data and process results toward perfect communication and improved performance for the entire system.

This way it allows easy transfer of information, user authentication processes, and log maintenance. Secure entry into the core engine is ensured by this module while also confirming who users are by recording how they interact with the system fully. Together with other modules like the Core Engine, the API integration module is vital because it facilitates the exchange of information as well as ensures strong systems at work.

The module that fetches videos is an important part since it bridges the gap between user prompts and the actual video content. This module interprets and processes queries from the front end interface to obtain information like frames and timestamps. The use of concurrency in the module ensures that many instances of video data are operated on at once thereby increasing its speed and ability to respond quickly. After processing videos, this module sends them to a core engine for storage or further analysis which makes the system effective and friendly to users.

IV. RESULTS AND DISCUSSION

Language Understanding Module (LLM) has shown its ability to extract important information from user input prompts. The case of the given prompt "Find the time between 12 AM and 3 PM when a blue SUV with the plate number MH09X4587 entered the parking area." illustrates how LLM can accurately identify and classify elements like color, vehicle type, and license plate number. This module uses natural language processing techniques to efficiently understand users questions and change them into structured JSON, which is further analyzed for appropriate responses.

Fig.2: LLM Output from the User Input Prompt

In Fig. 2 above, this diagram shows a visual representation of the extracted information in JSON output format. In such as an organization, features like color, cars's makes, as well as Registration numbers are arranged systematically, allowing easy access and understanding by other system parts. For example, It is used in structuring data for efficient handling by users. queries, resulting in smooth communication between users and machines. The JSON output demonstrates how unstructured Inputs are converted into actionable data by the system, thus enabling effective video content extraction and analysis.

Fig.3: API Routes

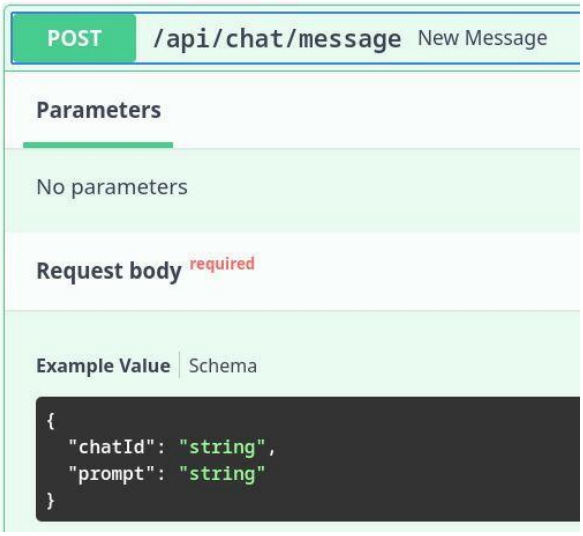


Fig.4: Request body of "/api/chat/message" that creates new message

Also, the API module has successfully completed important tasks. vital to the system, such as generating chat IDs and uploading videos. Fig. 3 shows that a new chat instance is started. by the POST request to the ”/api/chat/” endpoint through our system’s API implementation. Here, people can talk within the system. Afterward, users are required to upload video footage. pertaining to the chat. The process of uploading involves making a POST request to the ”/api/footage/upload” endpoint represented in Fig. 4. It receives files and then employs our model to find what objects there are in the footage uploaded. here. Thereafter, these objects and their timestamps get stored in a knowledgebase for further analysis.

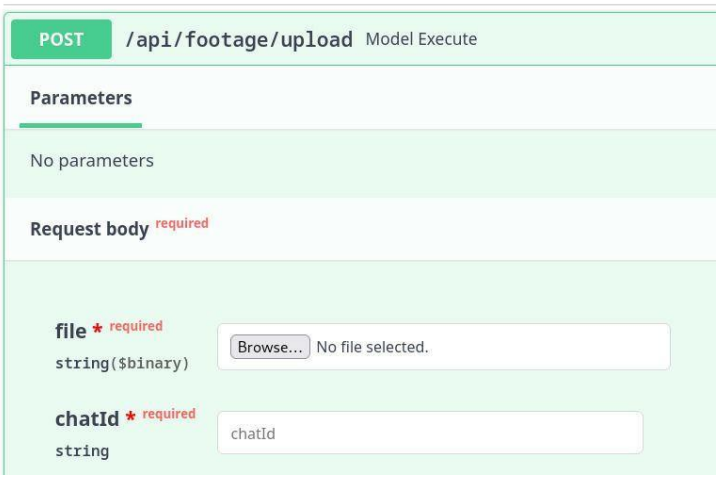


Fig.5: Request body of "/api/footage/upload" that uploads video and runs the model

A user may create new messages within the chat once. footage has been uploaded and processed through a POST request towards the ”/api/chat/message” endpoint. Fig. 5 explains how the request body looks when creating such a message. Moreover, any user could be able to fetch all chats just by doing GET requests on the”/api/chat” endpoints. Also, data about any given chat can be accessed through a GET request on ”/api/chat/?id” endpoint with chat ID as a parameter value entered into it as well. In addition, users can delete particular chats by sending DELETE requests via URLs ending with “id,” where “id” represents this specific chat’s ID again.

V. CONCLUSION

In conclusion, the AI-powered video analysis system signifies a major advancement in surveillance technology. Integrating Python, YoloV8, FastAPI, PostgreSQL, NextJS, TypeScript, and Tauri, the system provides a platform for video content extraction. The core engine, fueled by YoloV8 and Python, facilitates real time object detection, while the API module ensures secure communication. The Python based video fetcher enhances user interaction, and the user friendly interface, developed with NextJS, TypeScript, and Tauri, enables excellent use. The system promises a continuous monitoring, scalability, real time notifications, user training, authentication, and ethical guidelines. This positions the system as a responsible and user friendly solution, contributing to security, efficiency, and the application of AI in surveillance, among other areas.

VI. ACKNOWLEDGMENT

The progression of our work owes much to the invaluable support and guidance from various individuals. Our sincere thanks to the management of St. Joseph's College of Engineering and Technology for this opportunity. Special gratitude to Dr. V. P. Devassia, our Principal, for his unwavering support. We appreciate the valuable contributions of Dr. Joby P.P. and Prof. Kishore Sebastian. A special acknowledgment to our project guide, Prof. Smitha Jacob, and the entire staff of the Department of Computer Science and Engineering for their constant encouragement and support throughout this phase.

VII. REFERENCES

[1] You Only Look Once: Unified, Real-Time Object Detection by Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi

[2] A Review of Yolo Algorithm Developments by Peiyuan Jiang, Daji Ergu*, Fangyao Liu, Ying Cai, Bo Ma, The 8th International Conference on Information Technology and Quantitative Management (ITQM 2020 \& 2021)

[3] Research on Traffic Sign Detection Based on Improved YOLOv8 by Zhongjie Huang1, Lintao Li1, Gerd Christian Krizek2, Linhao Sun, Journal of Computer and Communications, 2023, 11, 226-232

[4] YOLO-PAI: Real-time handheld call behavior detection algorithm and embedded application by Zuopeng Zhao, Tianci Zheng, Kai Hao, Junjie Xu, Shuya Cui, Xiaofeng Liu, Guangming Zhao, Jie Zhou, Chen He

[5] YOLO-based Human Action Recognition and Localization by Shubham Shindea, Ashwin Kotharia, Vikram Gupta

[6] YOLO*C — Adding context improves YOLO performance by Goran Oreski

[7] Object detection in crowded scenes via joint prediction by Hong-hui Xu, Xin-qing Wang, Dong Wang, Bao-guo Duan, Ting Rui

[8] Faster-YOLO: An accurate and faster object detection method Yunhua by Yin, Huifang Li, Wei Fu

[9] A real-time object detection algorithm for video by Shengyu Lua, Beizhan Wanga, Hongji Wanga, Lihao Chenb, Ma Linjiana, Xiaoyan Zhangc

[10] What are large language models supposed to model? by Idan A. Blank Cognitive Sciences; Vol 27

[11] AccDecoder: Accelerated Decoding for Neural-enhanced Video Analytics by Tingting Yuan, Liang Mi, Weijun Wang†‡, Haipeng Dai, Xiaoming Fu; University of Gottingen, Germany; Nanjing University, China

[12] Video analytics using deep learning for crowd analysis by Md Roman Bhuiyan1, Junaidi Abdullah1, Noramiza Hashim1 & Fahmid Al Farid1; Multimedia Tools and Applications (2022) 81:27895–27922, https://doi.org/10.1007/s11042-022-12833-z

[13] Language as Queries for Referring Video Object Segmentation by Jiannan Wu, Yi Jiang2 Peize Sun, Zehuan Yuan, Ping Luo1, The University of Hong Kong, ByteDance, HKU-TCL Joint Research Centre for Artificial Intelligence

[14] Forensic Digital Analysis For CCTV Video Recording by Pria Sukamto, Ispandi, Arman Syah Putra, Nurul Aisyah, Rohmat Toufiq, International Journal Of Science, Technology & Management, ISSN: 2722-4015

[15] Multi-frame-based adversarial learning approach for video surveillance by Prashant W. Patil, Akshay Dudhane, Sachin Chaudhary, Subrahmanyam Murala; Pattern Recognition 122 (2022) 108350

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

Electronic Health Record Using Blockchain

Ms. Kesiya Johnson[1], Dr. Sarika S[2]
[1]Department of Computer Science, Naipunnya Institute of Management and Information Technology, Thrissur, Kerala, India
[2]Department of Computer Science and Engineering, Viswajyoti College of Engineering and Technology, Vazhakulam, Kerala, India

ABSTRACT
Blockchain technology has long been a subject of discussion in academic circles, and numerous businesses have benefited from it. Blockchain technology's decentralization, security, privacy, and other advantages make it incredibly advantageous for the healthcare industry. However, there are issues with data management, security, and integrity in electronic health records (EHRs). In this article, we investigate how blockchain technology might revolutionize EHR systems and perhaps address these issues. We outline a framework for integrating blockchain technology with electronic health records in the healthcare sector. Our proposed framework aims to do two things: first, it will apply blockchain technology to EHRs; second, it will protect electronic record storage by giving users of the framework granular access rules. Furthermore, this architecture addresses the issue of scalability that blockchain technology generally faces by utilizing off-chain record storage. This framework offers the scalable, secure, and essential blockchain-based solution to the EHR system.
Keywords: EHR, InterPlanatery File System (IPFS), Blockchain, Decentralisation.

I. INTRODUCTION

The primary benefits of technology progress lie in enhanced safety, user-friendliness, and other healthcare-related features. The recent development of technology has altered how we utilize and view the world, having an impact on every element of human existence. Similar to how technology has altered many other facets of life, it is also discovering novel approaches to enhance healthcare. Electronic medical records (EMR) and electronic health records (EHR) have made these advantages possible. Nonetheless, there are still certain problems with data integrity, user ownership, and security of medical records, among other things. Using a new technology, blockchain, could be the answer to these issues. This technology offers a safe, temperature-controlled platform for the storage of medical records and other data pertaining to health. Prior to the development of contemporary technology, medical records were kept on paper using a handwritten method in the health care industry. This paper-based system lacked organization, security, and efficiency. Due to the fact that the patient's medical record was duplicated and redundant in each of the facilities the patient visited, it also had to deal with this issue. An emerging trend in the healthcare sector was the integration of paper-based and electronic medical records (EMRs) through EHR systems.

Clinical notes and lab results were stored in these systems' many components. By reducing errors and facilitating information availability, they were meant to enhance patient safety. EHR systems were designed to replace paper-based health records with an efficient system that would transform the healthcare industry. They are considered to be a vital part of the healthcare sector since they greatly enhance its functionality. Organizing patient appointments, accounting and billing, lab testing, and electronic medical record keeping are some of these capabilities. These are available in many of the EHR systems used in the healthcare sector. Providing safe, dependable, and platform-transferable medical records is the primary objective.

II. LITERATURE REVIEW

Blockchain technology, already utilised in other industries, is set to be adopted in the healthcare industry. Numerous scholars have undertaken study in this area, with their findings concentrating on the feasibility of using blockchain technology into the healthcare business. They also discuss the benefits, dangers, concerns, and challenges associated with employing this technology. Some researchers have discussed the challenges of really putting this into reality on a larger scale. Some of the latest research are listed below:

- Wang and colleagues conducted a research into the usage of smart contracts in blockchain technology. **Error! Reference source not found.**The smart contracts and their operating framework are initially introduced. Operating systems and associated concepts. The authors also discuss how smart contracts may be used to the innovative concept of parallel blockchains. They explain that the justification for implementing blockchain smart contracts stems from the decentralisation afforded by the programming language code embedded in them. Upon introduction, foundations of smart contracts The author discussed the several blockchain layers that interact to keep the system working. Data, network, consensus, application contract, layer, and incentive.
- Vujii et al **Error! Reference source not found.** offered an overview of blockchain technologies, including bitcoin and Ethereum. According to the authors, the environment of information technology is always developing, and blockchain technology benefits information systems.They defined Bitcoin as a decentralised peer-to-peer network used for bitcoin transactions. Along with defining blockchain mining, they developed the proof-of-work consensus algorithm. The authors emphasise that blockchain scalability is a real issue, and that many solutions, such as SegWit and Lightning, Bitcoin Cash, and Bitcoin Gold, have been proposed to address it. The paper also outlined Ethereum and its dependencies, as well as the differences between the Ethereum and Bitcoin blockchains.Understand the scientific phrases and jargon relevant to your research.
- Gordon and Catalini's study **Error! Reference source not found.** looked at how blockchain technology might help the medical field. They came to the conclusion that the healthcare sector is dominated by hospitals, pharmaceutical companies, and other related parties. Data sharing was found to be the main argument in favour of employing blockchains in the healthcare industry. The healthcare sector needs to adapt four methods or components, according to this report, in order to use blockchain technology. Digital access rights, data accessibility, and expedited access to clinical information and patient identity are among the issues that can be handled. Both off-chain and on-chain data storage are covered. The study also discussed the challenges or roadblocks associated with applying blockchain technology, including the large volume of medical records, privacy issues, and security issues.

- Sahoo and Baruah **Error! Reference source not found.** created a scalable framework for blockchain technology using the Hadoop database. To solve the scaling issue with blockchain, they proposed fusing the scalability provided by the underlying Hadoop database with the decentralisation provided by blockchain technology. They employed a method to store blocks on the Hadoop database to improve the scalability of the blockchain technology. All required blockchain dependencies are present in the blockchain that is constructed using this framework. This research suggests using the Hadoop database system in conjunction with SHA3-256 hashing for blocks and transactions to overcome the blockchain platform's scalability problem. The coding language used for this architecture was Java.

III. BLOCKCHAIN TECHNOLOGY

This technique was provided by Nakamoto, the person who created the well-known cryptocurrency or digital currency known as bitcoin. Nakamoto used blockchain technology to solve the bitcoin double spending problem, but this innovative technology was soon being used for a number of additional purposes. The blockchain is an ever-expanding network of linked blocks that records new transactions on a regular basis. This platform uses a decentralised system that allows information to be spread and each dispersed piece of data data to be jointly owned **Error! Reference source not found.**. Blockchains, which contain batches of cryptographically secured hashed transactions, are managed via peer-to-peer networks. A new block is created on the blockchain network when a user sends a new transaction. Transactions are stored in blockchain blocks, which are sent to every node connected to the network[8]. Each transaction included in a block is disseminated to every node in the network. The verification process is facilitated by the fact that each node in the network has a copy of the whole blockchain. Before broadcasting a user transaction, every linked node verifies that it hasn't been altered in any way.The nodes add the block to their own copy of the blockchain if the verification is successful.Figure 1 clearly illustrates the Blockchain's architectural concept.The process of adding a block to the blockchain involves the nodes reaching a consensus on which blocks should be added and which shouldn't. In order to validate the transaction and ensure that the sender is an authorised network member, the connected nodes use a few well-known mechanisms. A node receives cryptocurrency if it successfully completes the validation process.This process of validating transactions is called mining, and the node that is carrying it out is known as the miner.The block is added to the blockchain after validation is complete.The transaction is finished once the entire validation process has been completed.

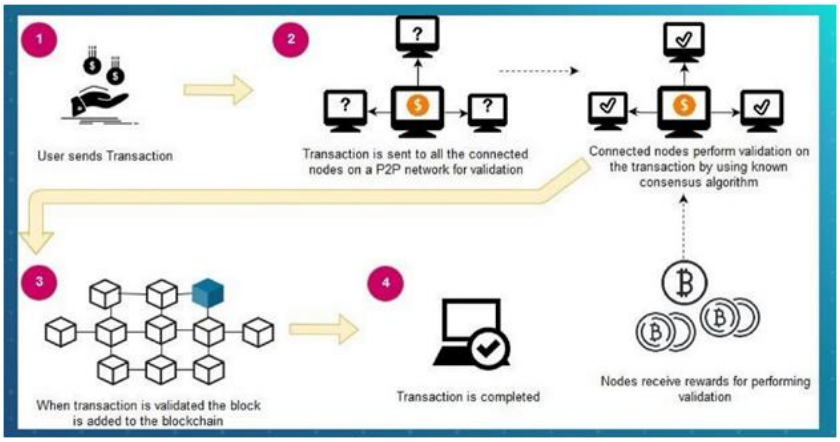


Fig 1. Blockchain architecture

I. KEY FEATURES OF BLOCKCHAIN

- **DECENTRALIZATION**

Instead of being centralized, data is shared throughout the network with blockchain. Because of this, information control is also dispersed and controlled by consensus determined by pooled input from all of the network's linked nodes. Several trustworthy entities currently oversee the data that was formerly concentrated at a single site.

- **DATA TRANSPARANCY**

Trust-based connections between entities are necessary for achieving data openness in any system. The required data or document must be protected.The blockchain distributes all stored data throughout the network, preventing it from being centralized in one place or controlled by a single node.

- **SECURITY AND PRIVACY**

The blockchain network's nodes are safe because of cryptographic procedures. The Secure Hashing Algorithm, or SHA, ensures data integrity through hashes, which provides security to the blockchain. Cryptographic hashes use strong one-way functions to prohibit data extraction and generate a checksum for digital data.The SHA-256 cryptographic algorithm is used to hash the hashes that are kept on the blocks[7]. Blockchain is thus a decentralized platform that is safe because of cryptographic methods, which makes it a great substitute for apps that have to protect user privacy.

IV. SOFTWARE PLATFORMS

❖ **ETHEREUM**

The concept behind Ethereum was to develop a trustless smart contract platform that would be open-source and would also hold the characteristic of programmable blockchain. Ethereum was initially released in the year 2015. The peer-to-peer networking that enables this technology distributed is also shared. This network also utilizes Ethers, a proprietary cryptocurrency. It is possible to transfer this cryptocurrency between accounts linked to theEthereum network.

❖ **INFORMATION TRANSACTION**

The interaction between an external entity and Ethereum is known as a transaction. External users may use it to update the status of records or other data kept on the Ethereum blockchain network. The following components make up an Ethereum transaction.

i. From - the 20-byte address of the message's sender.
ii. To - the recipient of the message, who likewise has a 20-byte address.
iii. Value - the amount of money sent from sender to recipient .The message being sent to the recipient is contained in the data, which is optional.
iv. Gas - The sender must pay a price for carrying out each transaction on the Ethereum blockchain. This fee is referred to as Gas.

❖ SMART CONTRACT

The piece of code that is used to carry out any task on the blockchain is known as a smart contract. The users' transactions are sent when this piece of code is run. They immediately run on the blockchain, making them impervious to hacking and modifications of any type. Programmers can utilize smart contracts, which often use the Solidity programming language, to code any kind of activity they wish to carry out on the blockchain. Programmers can use EVM bytecode, which is discussed in more depth in the section below, to compile the required operations once they have been created in code. After compilation, they might then be executed and made available on the Ethereum blockchain. Python and JavaScript are programming languages that are wrapped by Ethereum's Solidity language to build code for smart contracts.

• ETHEREUM VIRTUAL MACHINE (EVM)

One of the Ethereum platform's main advantages is its programmable blockchain. It gives users the option to develop custom applications that work with Ethereum. Distributed Applications (DApps) are the name given to the applications created on this platform. They include a number of protocols that are bundled up to make a DApp platform. These DApps include smart contracts with user-defined code that carry out specific application task definitions.As a result, the smart contract-based applications are actually running on EVM.

• INTERPLANATERY FILE SYSTEM (IPFS)

A peer-to-peer network is utilized by an IPFS technology for data storage. IPFS provides safe data storage since it guards against data manipulation. IPFS employs a cryptographic identity to protect the data from manipulation, since changing the identification is the only way to change anything recorded on the system. Every data file stored on IPFS contains an encrypted hash value. It serves as a unique means of identifying data files that are stored on the IPFS.An IPFS protocol is a preferable solution for storing sensitive and crucial data because of its secure storage mechanism.**Error! Reference source not found.** A decentralized application could cache the generated cryptographic hash to reduce the laborious computations on the blockchain.

V. SYSTEM DESIGN AND ARCHITECTURE

The section devoted to related work includes the work done in the healthcare industry using blockchain technology. As said before, they provided particular solutions to address the typical problems with blockchain technology. The majority of the research that were taken into account concerned blockchain technology-related scaling and data sharing difficulties. Unlike existing approaches, our proposed framework leverages IPFS's off-chain scaling technique to solve the scalability problem. Furthermore, the proposed framework is implemented in its entirety using Ethereum. This essay has already discussed Ethereum and its dependencies.

• SYSTEM DESIGN

The most crucial and significant component is system design because it is employed in system development. The modules, architecture, and other components that make up the framework of the entire system are included in this section. As previously stated, the proposed architecture aims to build a Decentralised-blockchain-based platform for electronic health data that is temper-proof, trustworthy, and private.Our system would continue to function if these modules were integrated. These entities or modules have additional notions

that are important to comprehend and are described in the sections that follow.Users of the suggested framework include medical professionals, nurses, administrators, and patients. They received granular access because they ought to have different levels of system authority.

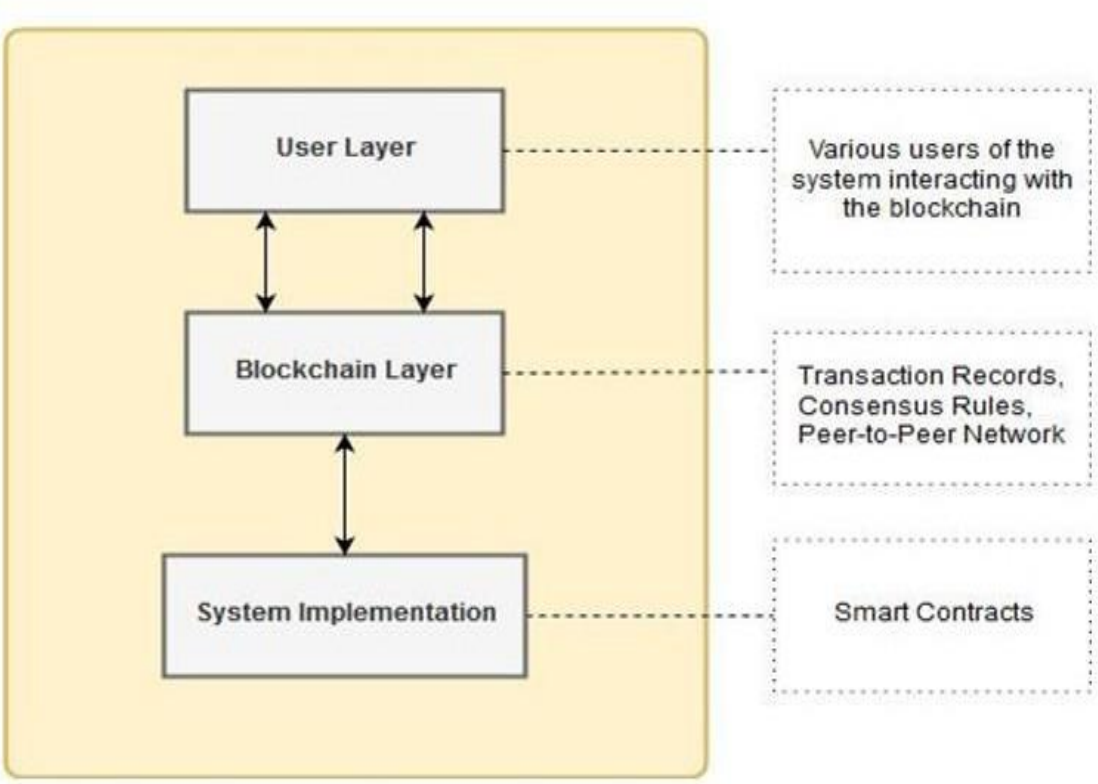


Fig 2 : Different levels of System design.

a) USER LAYER

A person who effectively utilizes a system's resources is referred to as a user of the system. A user can be recognized on the system thanks to his various roles and features.Users of this system may include administrative personnel, medical professionals, and patients. These users' primary duties would be to interact with the system and carry out simple operations including creating, reading, updating, and deleting medical records. The functionality of this system would be accessible to users through a browser, which we will refer to technically as a "DApp browser" because it houses the GUI (Graphical User Interface) of the DApp, or our suggested system framework.All of the functions that a certain user could access are contained in the GUI. This GUI allows users to interact with the blockchain layer of the system in accordance with their allocated roles.

b) BLOCKCHAIN LAYER

- Blockchain Resources : A third party can edit a record or other item of data that has been uloaded to the networkby using a transaction on the Ethereum blockchain.Since these transactions involve data that users can share with other users or just store for later usethe Ethereum blockchain treats them as assets.
- Governance Rules : Blockchain technology generally adheres to certain consensus criteria for the completion and computation of its transactions. The blockchain must be kept secure and temper-proof for this reason, thus some consensus techniques are required. The Proof of Work (PoW) consensus

method is used by the Ethereum blockchain to ensure that blockchain governance is upheld in a trustworthy way by agreement from all of the trusted nodes connected to the blockchain network.

- Network : The peer-to-peer network is used by the Ethereum blockchain. In this network, every node has a peer connection.There is no central node since no single node is in charge of managing every aspect of the network. This network was utilized since a distributed platform was intended, not a centralized one. Therefore, using a network where all linked nodes have equal status and privileges was the best thing this technology could have done.

c) TRANSACTIONS

- ADD RECORDS : Medical records for patients would be created by adding records to the DApp. It includes the ID, name, co-morbidity, blood type, and IPFS hash fields. Along with the IPFS hash of the file uploaded containing the patient's test results or other medical information, the patient's basic medical records are stored.
- UPDATE RECORDS : The patient's medical records would be updated. The IPFS hash cannot be changed by this; only the patient's fundamental information may. The IPFS hash cannot be updated to ensure record security.
- VIEW RECORDS : View records would allow the user to access a patient's saved DApp medical records. Both patients and doctors use the view records feature. By validating that the patient only reads his own medical records, the system allows the patient to examine his records. The patient's public account address is used by the system for this reason to guarantee that only the pertinent medical records are displayed to the patient.
- DELETE RECORDS : The ability to erase records would allow the user to do so for any patient. The users in this case would be the doctors, who have the authority to remove any patient information from the blockchain.
- PERMIT: Permit access for each of the aforementioned transactions; only a certain user needs access to them.

VI. IMPLEMENTATION RESULTS AND ANALYSIS

➢ AVERAGE TIME OF EXECUTION

It takes longer to execute the more transactions there are. These transactions support the different smart contract functionalities, the algorithm for which is defined in Section V. The system's Assign Roles, Add Patient Records, and View Patient Records functions would run in 18.29 seconds, 1 minute, 48 seconds, and 50 seconds, respectively, when just one user is logged in. The time would increase if 100 users were using the system at once.

➢ THROUGHPUT

Utilizing JMeter, we created a simulation of 100 to 500 people utilizing the system and carrying out its various activities over a period of 10 to 35. JMeter measures throughput in data/time, or KB/sec, units. We tested the system's performance when doing the trials while simulating the above-mentioned number of users. These simulations are carried out inside the suggested framework, and throughput is analyzed at the end.This

experiment showed that when the number of users and requests increased, the system's throughput increased dramatically in a linear fashion. The effectiveness of the proposed structure is demonstrated by this linear development in throughput

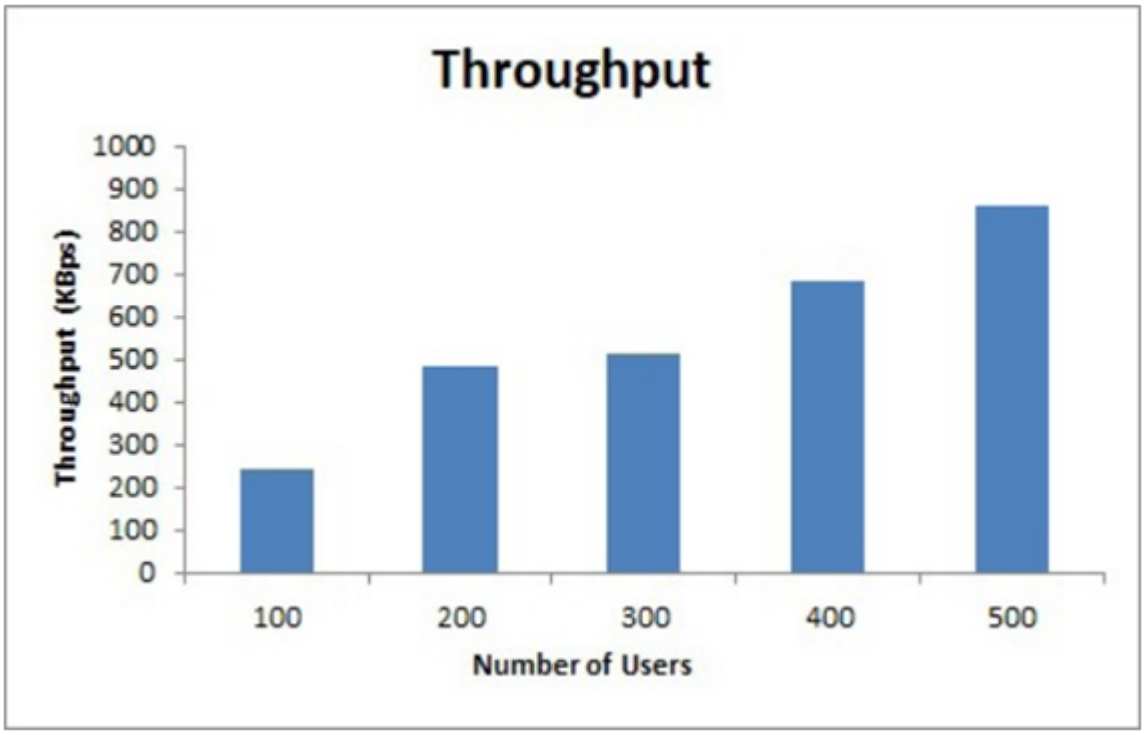


Fig 3 : Throughput analysis

- **MEDIAN LATENCY**

The temporal lag that exists between a system component sending a request and any other system component producing a response is known as latency. What separates these two actions is their latency. Here, we evaluated the recommended framework's average latency using JMeter. We tested the latency of the proposed framework by simulating the number of users using JMeter.JMeter uses milliseconds as a measure of delay.

VII. COMPARISON OF PROPOSED FRAMEWORK WITH RELATED WORK

We also go over some of the parameters that comprise our framework and are employed to contrast it with previous studies in the area. It is ensured that these requirements are included in the framework, but it is also taken into account that doing so won't compromise system security or privacy. Each of the characteristics listed below addresses security and privacy in relation to this.

- **SCALABILITY**

Simplifying the meaning, scalability refers to an information system's ability to function properly when its storage volume increases or decreases. Blockchain technology's scalability issue calls for a long-term solution. The volume and size of data on the blockchain are increasing. Our proposed system used the off-chain storage technique because the patient's data kept on the blockchain includes the IPFS hash in addition to the patient's fundamental information. The off-chain scaling method employed in our suggested system architecture is this one. Since a significant portion of patient medical records are no longer stored on the blockchain, the scalability

issue has been fixed. Transactions can now be finished faster on the blockchain because the majority of the data stored there has decreased. As was already said, IPFS makes use of cryptographic hashes that are decentralized and stored via a peer-to-peer network. This ensures that while the scalability problem is being resolved, the framework's security remains unaffected.

➢ CONTENT-ADDRESSABLE STORAGE

The off-chain storage component of IPFS used in the suggested framework is referred to as content-addressable storage.The IPFS is used to store the patient's sensitive record, which makes sure that a hash of the record is created. The blockchain now contains that hash, which the doctors and patients can access as needed. The security of the data stored on IPFS is ensured by the cryptographically secure hash that is generated. Additionally, this guarantees security in our suggested framework.

➢ INTEGRITY

The degree of dependability, stability, and resilience of the information storage system determines the integrity of the system. This system maintains correct data that hasn't been tampered with in any way. Furthermore, the information is only accessible to patients and clinicians who are associated parties. Because they do not have access to the smart contract, users of the system and other parties are not allowed to change it. Utilizing the access rules, which ensure that patient medical records are kept private and unreadable, allows for this. Additionally, by storing records on IPFS, the security of the patients' medical records is guaranteed.

➢ ACCESS CONTROL

The Role-based access mechanism used by this framework ensures that each system entity is given a role.The system would not be accessible to any third parties that are not allowed to use it. This system offers two main forms of security. First, blockchain technology is safe by nature and adheres to specific standards to protect itself from outside assaults. Furthermore, role-based access is employed by our system, limiting user access to its features to those who have been allocated particular roles. As a result, our solution would ensure not only the safety of patient records but also the management of access for companies that are associated with them.

➢ INFORMATION CONFIDENTIALITY

To protect the privacy of the patients' data, the patient medical records stored on the blockchain should be protected from access by outside parties. The patient's data includes crucial information on the patient, including blood type, records, lab results, X-ray reports, MRI results, and numerous other relevant results and reports.The hospital depends on all of this information in addition to the patients. Smart contracts are a very helpful part of this system since they provide confidence, accuracy, and precision in the transactions that are completed. The records that the system saves and retrieves are only accessible to reliable persons. Any attempt to access the system by an untrusted outsider is denied.

VIII. CONCLUSION

This research focused on how blockchain technology may be applied to electronic health records in the healthcare business. Despite the growth of the healthcare business and technical advancements in EHR systems,

there were several issues that this new technology, or blockchain, was able to fix. Our architecture combines granular access constraints and secure record storage. It creates an easier method for people to utilise. The system makes advantage of IPFS's off-chain storage capability, and the framework also includes advice on how to ensure that the data storage problem is addressed. Furthermore, the system benefits from role-based access, which allows only connected and trustworthy persons access to medical records. This also addresses the EHR system's information asymmetry issue.

➢ FUTURE WORKS

We intend to use the current framework to identify session medjacking.The term ”Session medjacking” refers to the act of hacking a medical device with the intention of endangering or threatening a patient.It may bring the patients to the point of death.By seeing this, we can stop third parties from misusing information or posing dangers. Additionally, a patient medical insurance module and an appointment module would need to be implemented.

IX. REFERENCES

[1] M. S. Sahoo and P. K. Baruah, ‘‘HBasechainDB—A scalable blockchain framework on Hadoop ecosystem,’’ in Supercomputing Frontiers. 2018, pp. 18–29.

[2] M. Hochman, ‘‘Electronic health records: A ‘‘Quadruple win,’’ a ‘‘quadruple failure,’’ or simply time for a reboot?’’ J. Gen. Int. Med., vol. 33, no. 4, pp. 397–399, Apr. 2018. B.

[3] S. Wang, Y. Yuan, X. Wang, J. Li, R. Qin, and F.-Y. Wang, ‘‘An overview of smart contract: Architecture, applications, and future trends,’’ in Proc.IEEE Intell. Vehicles Symp. (IV), Jun. 2018, pp. 108–113.

[4] D. Vujičić, D. Jagodić, and S. Randić, ‘‘Blockchain technology, bitcoin,and Ethereum: A brief overview,’’ in Proc. 17th Int. Symp. INFOTEHJAHORINA (INFOTEH), Mar. 2018, pp. 1–6.

[5] W. J. Gordon and C. Catalini, ‘‘Blockchain technology for healthcare:Facilitating the transition to patient- driven interoperability,’’ Comput.Struct. Biotechnol. J., vol. 16, pp. 224–230, Jan. 2018.

[6] M. Reisman, ‘‘EHRs: The challenge of making electronic data usable and interoperable.,’’ PT, vol. 42, no. 9, pp. 572–575, Sep. 2017.

[7] S. T. Argaw, N. E. Bempong, B. Eshaya-Chauvin, and A. Flahault,‘‘The state of research on cyberattacks against hospitals and available best practice recommendations: A scoping review,’’ BMC Med. Inform.Decis.Making, vol. 19, no. 1, p. 10, Dec. 2019.

[8] M. Niranjanamurthy, K. Kumar S, A. Saha, and D. D. Chahar, ‘‘Comparative study on performance testing with jmeter,’’ Int. J. Adv. Res. Comput.Commun. Eng., vol. 5, no. 2, pp. 70–76, 2016.

[9] P. Zhang, J. White, D. C. Schmidt, G. Lenz, and S. T. Rosenbloom,‘‘FHIRChain: Applying blockchain to securely and scalably share clinical data,’’ Comput. Struct. Biotechnol. J., vol. 16, pp. 267–278,Jul. 2018.

[10] J. Eberhardt and S. Tai, ‘‘On or off the blockchain? Insights on offchaining computation and data,’’ in Proc. Eur. Conf. Service-Oriented Cloud Comput., Oct. 2014, pp. 11–45.

[11] Zheng, Z., Xie, S., Dai, H. N., Chen, X., & Wang, H. (2017). Blockchain challenges and opportunities: A survey. International Journal of Web and Grid Services, 14(4), 352-375. This paper presents a survey of the challenges and opportunities associated with blockchain technology.

[12] Swan, M. (2017). Blockchain thinking: The brain as a decentralized autonomous corporation. IEEE Technology and Society Magazine, 36(2), 6-13. This paper explores the concept of blockchain thinking and its implications for decentralized systems.

[13] Zheng, Z., Xie, S., Dai, H. N., Chen, X., & Wang, H. (2018). Blockchain challenges and opportunities: A survey. International Journal of Web and Grid Services, 14(4), 352-375. This paper presents a survey of the challenges and opportunities associated with blockchain technology.

[14] Crosby, M., Pattanayak, P., Verma, S., & Kalyanaraman, V. (2016). Blockchain technology: Beyond bitcoin. Applied Innovation, 2(6-10), 71-81. This paper explores various applications of blockchain technology beyond cryptocurrencies like Bitcoin.

[15] Christidis, K., & Devetsikiotis, M. (2016). Blockchains and smart contracts for the internet of things. IEEE Access, 4, 2292-2303. This paper discusses the integration of blockchain technology and smart contracts with the Internet of Things (IoT).

[16] Zheng, Z., Xie, S., Dai, H. N., Chen, X., & Wang, H. (2017). An overview of blockchain technology: Architecture, consensus, and future trends. In 2017 IEEE international congress on big data (pp. 557-564). IEEE. This paper provides an overview of blockchain technology, including its architecture, consensus mechanisms, and future trends.

[17] Yli-Huumo, J., Ko, D., Choi, S., Park, S., & Smolander, K. (2016). Where is current research on blockchain technology? A systematic review. PloS one, 11(10), e0163477. This paper presents a systematic review of the current research landscape surrounding blockchain technology.

[18] Tapscott, D., & Tapscott, A. (2016). Blockchain revolution: How the technology behind bitcoin is changing money, business, and the world. Penguin.

[19] Swan, M. (2015). Blockchain: Blueprint for a New Economy. O'Reilly Media, Inc. This book provides an in-depth exploration of blockchain technology and its potential applications.

[20] Swan, M. (2015). Blockchain: Blueprint for a New Economy. O'Reilly Media, Inc. This book provides an in-depth exploration of blockchain technology and its potential applications.

[21] Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System. This is the original whitepaper that introduced the concept of Bitcoin and the blockchain technology it is built upon.

[22] Xia, Q., Sifah, E. B., Smahi, A., Amofa, S., & Zhang, X. (2017). BBDS: Blockchain-based data sharing for electronic medical records in cloud environments. Information, 8(2), 44. This paper proposes a blockchain-based data sharing system for electronic medical records in cloud environments.

[23] Yue, X., Wang, H., Jin, D., Li, M., & Jiang, W. (2016). Healthcare data gateways: Found healthcare intelligence on blockchain with novel privacy risk control. Journal of medical systems, 40(10), 218. This paper introduces healthcare data gateways, which leverage blockchain technology to ensure secure and privacy-preserving sharing of healthcare data.

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

Sobot-Solar based Beach Cleaning Robot

Aiswarya Jose, Ann Mariya Shaji, Isaac George, Sharon Merin Sabu
St. Joseph's college of Engineering and Technology, Palai Kerala, India

ABSTRACT

Beaches are not only popular tourist destinations but are also an asset to the environment. In these days, the pollution on beaches are increasing day by day. The major wastes that are accumulated on the beach shore includes broken glass pieces, medical waste, jellyfish, plastic bottles and bags, rusted metal, etc. These waste ends up in the sea if they are not collected properly and manual collection of this waste will result in health problems for beach workers. The Sobot is a solar based beach cleaning robot that collects waste with minimum human intervention. It uses a combination of sensors, such as ultrasonic sensors, IR sensors, cameras, and GPS to detect obstacles and navigate its path respectively. We implement the image detection and processing using the micro controller ESP 32 which has an inbuilt camera module which captures the images. The Sobot operates on the sandy shores, harnessing the power of the sun through integrated solar panels. Our aim is to design and develop an affordable, easily portable and environmental friendly machine that solves the issue of beach pollution.

Keywords: Machine, manual collection, waste collection, sensors, GPS, solar based.

I. INTRODUCTION

Oceans account for 70 percent of the surface of Earth and play a pivotal role in the health of human beings. According to the National Oceanic and Atmospheric Administration (NOAA), billions of pounds of trash and other pollutants enter our oceans every year. This has become an issue not for our country but also for the entire world. Proper disposal of waste is crucial, for maintaining hygiene, cleanliness and overall environmental well being.

To tackle this situation we came up with the idea of Sobot, the beach cleaning robot is a piece of remote controlled equipment designed specifically for cleaning beaches. It has the ability to operate on dry terrain, effectively collecting trash and other forms of waste. The robot uses image processing algorithms to detect various objects on the beach and locate them within the robot's field of view.

In this paper we will explore the design, operation, and potential effects on beach laborers, highlighting its contribution to cleaner and healthier beaches while furthering the goals of environmental conservation and sustainable growth.

II. METHODS AND MATERIAL

Sobot is a four-wheeled skid-steer drive powered by two motors on front wheel, with electrical power provided by a battery pack inside the vehicle using solar panel. The waste detection is done using micro controller ESP32-CAM. The ESP32-CAM has a camera module which captures the image of object. After capturing the image of the object, the frames are transmitted to the cloud via ESP32-CAM. After comparing the captured object with the input database from the cloud, if the object is recognized as plastic bottle, then the wheel motor stops and initiates the arm motor which collects the waste. If the object is not recognized as plastic bottle, the image is ignored and it looks for upcoming images. The storage tank of the robot has a capacity of 10 litres, and after reaching the limit, the IR sensor gives the input to the control station, and the robot moves to the base station and returns back after emptying the tank. The robot uses a mesh like design for waste collection. The robot's design is eco-friendly, and it aims to help clean beaches of waste and junk.

III. RESULTS AND DISCUSSION

The development of a solar-based beach cleaning robot represents a promising step towards addressing the challenge of beach pollution in an environmentally sustainable manner. Through the integration of solar power, the robot offers an eco-friendly solution that reduces reliance on conventional energy sources and minimizes carbon emissions. Moreover, the autonomous operation of the robot enables efficient and systematic cleaning of beaches, effectively removing debris and waste while minimizing human intervention. The use of robotics in beach cleaning not only enhances the effectiveness of clean-up efforts but also serves as a demonstration of technological innovation in environmental conservation. Additionally, the utilization of renewable energy sources aligns with the global trend towards sustainability and contributes to the promotion of clean energy technologies. Overall, the solar-based beach cleaning robot holds great promise in mitigating beach pollution and preserving the natural beauty of coastal ecosystems.

IV. CONCLUSION

In conclusion, the idea of solar-based beach cleaning robot represent has a promising and innovative solution to the challenges of maintaining pristine coastal environments. Sobot offer a sustainable and efficient way to address the issue of beach pollution, combining renewable energy technology, advanced robotics, and environmental conservation. With their ability to autonomously navigate diverse beach terrains, identify and remove debris, and operate under the power of the sun, this robot hold significant potential for local governments, tourist resorts, environmental organizations, and research institutions.

As we look to the future, the continued research and development in this field will further enhance the capabilities and effectiveness of solar-based beach cleaning robots. With a focus on sustainability, efficiency, and ecological impact, these robots represent a compelling example of how cuttingedge technology can address real-world environmental issues, offering hope for a cleaner and healthier coastal environment for generations to come.

V. REFERENCES

[1] H. Ebrahim, W. Sheikh, A. Saeed, "Design and analysis of sustainable beach cleaner", 3C Technology. Glosses of innovation.

[2] N. Bolong, I. Saad, M. Amran Madlan, "Manufacturing of Beach Cleaning Machine at University Malaysia Sabah (UMS) Prototype Design and Analysis", Transactions on Science and Technology, Volume: 08, No. 3-2, 281 - 289, 02 Nov 2021.

[3] Prof. J. Shelke, B. Bhakare, K. Lute, A. Pateshwari, H. Khodiyar, "Beach cleaning system and surface cleaning system", International Research Journal of Modernization in Engineering Technology and Science, Volume:02, Issue:06, e-ISSN: 2582-5208, June 2020.

[4] V. Mepani, H. Patel, Vataliya Mohil, Prof. R. Sahu, "Design and Fabrication of Beach Sand Cleaning Machine", International Research Journal of Engineering and Technology, Volume: 07, Issue: 02, e-ISSN: 2395-0056, p-ISSN: 2395-0072, Feb 2020.

[5] A. O. Qasim, A. A. Varghese, A. Sarkar, and S. Justus, "Modular Type Beach Cleaning Robot 'Clean-B,'" Proc. - Int. Conf. Augment. Intell. Sustain. Syst. ICAISS 2022, pp. 1231–1234, 2022, doi: 10.1109/ICAISS55157.2022.10010711.

[6] D. Varghese and A. Mohan, "Binman: An Autonomous Beach Cleaning Robot," MysuruCon 2022 - 2022 IEEE 2nd Mysore Sub Sect. Int. Conf., pp. 1–5, 2022, doi: 10.1109/MysuruCon55714.2022.9972499.

[7] Chandra, S. S., Kulshreshtha, M., Randhawa, P. (2021). A Review of Trash Collecting and Cleaning Robots. 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), ICRITO 2021, September 2021. https://doi.org/10.1109/ICRITO51393.2021.9596551

[8] J. Shalini Priya, K. T. Balaji, S. Thangappan, and G. Yuva Sudhakaran, "Beach Cleaning Bot Based on Region Monitoring," 8th Int. Conf. Comput. Power, Energy, Inf. Commun. ICCPEIC 2019, pp. 1–4, 2019, doi: 10.1109/ICCPEIC45300.2019.9082368.

[9] H. M. R. G. Herath, "Design and Implement of Beach Cleaning Robot Design and Implement of Beach Cleaning Robot," no. August, 2022.

[10] V. Dhole, O. Doke, A. Kakade, S. Teradale, Prof. R. Patil, "Design and fabrication of beach cleaning machine", International Research Journal of Engineering and Technology (IRJET), e-ISSN: 2395-0056, Volume: 06 Issue: 04, Apr 2019.

[11] H. Sukri, A. Dafid, and A. Basuki, "Design of Garbage Collection Robots In Tourism Area (Beach) With Artificial Neural Network Method," Proceeding - IEEE 8th Inf. Technol. Int. Semin. ITIS 2022, pp. 235–240, 2022, doi: 10.1109/ITIS57155.2022.10010038.

[12] C. Balasuthagar, D. Shanmugam, K. Vigneshwaran, "Design and fabri cation of beach cleaning machine", IOP Conf. Series: Materials Science and Engineering 912 (2020) 022048 IOP Publishing doi:10.1088/1757- 899X/912/2/022048, 2020.

[13] Prathamesh Jangam, Sneha jangam, Rutuja kadu, Mubbashir Kazi, Prof. Sanobar Shaikh, "Beach Cleaning Robot", International Research Journal of Engineering and Technology (IRJET), ISSN: 2349-6002, Volume:6 Issue:12, May 2020.

[14] Ramamoorthi R, Ramachandran N, Nikiles PD, Jayasurya R, Natheesh MD, Nithin K Biju, "Design and fabrication of beach cleaning ma chine", International Journal of Innovative Technology and Exploring Engineering, ISSN:2278-3075, Volume: 08, Issue:12, Oct 2019.

[15] C. P. Le, A. Q. Pham, H. M. La, and D. Feil-Seifer, "A Multi-Robotic System for Environmental Dirt Cleaning," Proc. 2020 IEEE/SICE Int. Symp. Syst. Integr. SII 2020, pp. 1294–1299, 2020, doi: 10.1109/SII46433.2020.9026295.

[16] M.Bhavani, S.Kalaiselvan, S.Jagan, S.Gopinath, "Semi-Automated Wireless Beach Cleaning Robot Vehicle", International Journal of Re cent Technology and Engineering (IJRTE), ISSN: 2277- 3878, Volume:8 Issue:1S2, May 2019.

[17] Dr. F B Sayyad, Dr. Md. Imran Ansari, Dr. S F Sayyad, "Design and Development of Beach Cleaning Machine", International Journal for Research in Applied Science and Engineering Technology, ISSN: 2321- 9653, Volume:07, Issue:06, June 2019.

[18] T. Subba Reddy, P. Satya Priyanka, L. Himaja, K. Sravani, N. Mounika, "Design and Fabrication of Beach Dust Collector", Research and Development in Machine Design, Volume:03,Issue:03,DOI:http://doi.org/10.5281/zenodo.4043052, Oct 2020.

[19] R Praveen, L Prabhu, P Premjith, Adarsh. K. Mohan, Ajayraj, "Design experimental of RF controlled beach cleaner robotic vehicle", IOP Conf. Series: Materials Science and Engineering 993 (2020) 012030 IOP Publishing doi:10.1088/1757-899X/993/1/012030, 2020.

[20] C. Zhao and Y. Liu, "The Yellow River Walker Beach Garbage Robot," Proc. - 2022 4th Int. Conf. Artif. Intell. Adv. Manuf. AIAM 2022, no. 2, pp. 768–772, 2022, doi: 10.1109/AIAM57466.2022.00155.

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

Enhancing Autonomous Driving Through Real-Time Steering Angle Prediction with Convolutional Neural Networks

Mr. Amalraj Joseph*[1], Dr. Santhosh Kottam[2]
*[1]Post Graduate Student, Department of Computer Applications, Federal Institute of Science and Technology, Angamaly, Kerala, India
[2]Professor, Department of Computer Applications, Federal Institute of Science and Technology, Angamaly, Kerala, India

ABSTRACT
The self-driving car's ability to handle curved roads is a significant breakthrough in autonomous driving research, holding promise for the development of real-world autonomous vehicles. This research contributes to autonomous vehicle technology by using Convolutional Neural Networks (CNNs) to predict steering angles in real time based on images from a mounted camera on the vehicle. The research utilizes a carefully chosen dataset from the AirSim simulator to train and assess the CNN model's performance in different driving situations, including complex tasks like navigating curves and turns. The results emphasize the potential of CNNs to enhance the accuracy of steering angle prediction, marking a notable accomplishment in the field of autonomous driving research. These findings have practical applications in the development of real-world autonomous vehicles.

Keywords- Convolutional Neural Networks, AirSim simulator, self-driving cars, classification

I. INTRODUCTION

Self-driving cars present a multitude of advantages that have the potential to revolutionize transportation. One key benefit is the capacity to enhance road safety by eliminating human error, a significant factor in accidents. Equipped with advanced sensors, cameras, and artificial intelligence, self-driving cars can perceive their surroundings, anticipate hazards, and react swiftly, ensuring safer roads. This technology aims to enhance safety, alleviate traffic congestion, and provide increased mobility for non-drivers. Additionally, these vehicles can improve traffic efficiency through real-time data sharing and optimized route planning, thereby reducing congestion, minimizing traffic jams, and saving both time and fuel. Another advantage is improved accessibility and mobility, which benefits individuals with physical disabilities, the elderly, and those facing mobility challenges. The reliable transportation options provided by self-driving cars enable easy commuting, participation in activities, and the pursuit of an active lifestyle. Moreover, self-driving cars enhance passenger productivity as driving tasks are automated. Passengers can utilize travel time for work, relaxation, or leisure activities, resulting in increased overall productivity and efficiency. Self-driving cars have the potential to

revolutionize transportation by offering increased safety, optimized traffic flow, improved accessibility, and enhanced productivity [4].

Convolutional Neural Networks (CNNs) are widely used artificial neural networks, particularly in computer vision. They excel at recognizing patterns and features in image data, making them ideal for tasks such as image classification and object recognition. CNNs are also well-suited for self-driving cars, as they can analyse real-time image data from onboard cameras to predict driving parameters, including the steering angle. The incorporation of CNNs into autonomous vehicles is a significant milestone, enabling cars to interpret their surroundings similarly to humans. Researchers and engineers continually strive to enhance the efficiency of these models, employing CNNs and other machine-learning techniques to further improve the performance and safety of self-driving cars. Simulated environments, such as the AirSim simulator, provide a valuable platform for testing and validating CNN-based models in diverse driving scenarios, which is crucial for developing real-world autonomous vehicles. These advancements in CNN-based self-driving car technology are expected to have a profound impact on future transportation [5][6].

II. BACKGROUND WORKS

Convolutional Neural Networks (CNNs) play a critical role in the development of self-driving cars, providing essential capabilities to perceive and comprehend the surrounding environment. Specifically designed for processing visual data, such as images and videos, CNNs are deep learning models employed for tasks like object detection, recognition, and tracking within the context of autonomous vehicles. By extracting significant features from visual input, CNNs enable accurate identification and classification of objects such as pedestrians, vehicles, traffic signs, and obstacles. Training CNNs on annotated datasets allows self-driving systems to learn complex patterns and understand contextual information, facilitating informed decision-making based on detected objects and their behaviour. Moreover, CNNs enable depth estimation and 3D mapping, providing crucial information for precise object localization and spatial understanding. This aids in enhancing safety by enabling real-time detection of potential hazards, proactive accident prevention, and prediction of other road users' behaviour. The continuous advancement of CNN algorithms and architectures holds promise for further improving the performance and efficiency of self-driving car systems, thus driving the broader adoption of this transformative technology.

The research work [1] Introducing a novel approach to developing a virtual self-driving car system, the study concentrates on enhancing Convolutional Neural Networks (CNNs) to improve the perception and understanding of the environment. The authors present their research findings and methodologies, offering insights into the training process and evaluation metrics employed. This work contributes to advancing the field of self-driving cars by demonstrating the effectiveness of the proposed improved CNN model in accurately detecting and classifying objects, ultimately facilitating safer and more reliable autonomous driving.

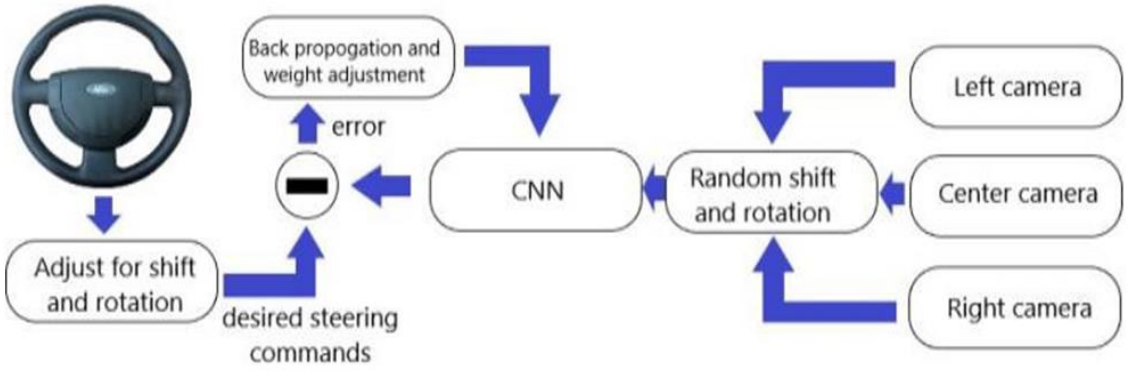


Figure 1: Architecture Diagram of Self Driving Car

The second case study [2] concentrates on simulating highly automated vehicle technology through the utilization of Convolutional Neural Networks (CNNs). The authors explore the application of CNNs in the context of self-driving cars, offering insights into the simulation process. This research contributes to advancing the field of self-driving cars by leveraging CNNs for simulation, thereby furthering the understanding of highly automated vehicle technology.

The next study [3] focuses on the utilization of Convolutional Neural Networks (CNNs) for end-to-end learning in steering angle prediction for autonomous electric vehicles. The authors, provide insights into the implementation of CNNs and their application in steering angle prediction. The research contributes to the advancement of autonomous vehicle technology by leveraging CNNs for accurate and efficient steering angle estimation, thereby enhancing the safety and performance of autonomous electric vehicles.

III. PRELIMINARIES

A. Convolutional Neural Network (CNN)

A Convolutional Neural Network (CNN) is a popular deep learning architecture widely used in Computer Vision tasks. Computer vision, a field of Artificial Intelligence, enables machines to comprehend and interpret visual data, such as images. Neural Networks, particularly CNNs, demonstrate excellent performance in various machine-learning applications involving diverse datasets, including images, audio, and text. CNNs are specifically designed for image classification tasks, while Recurrent Neural Networks, specifically LSTM, are suitable for sequence prediction.

1) Convolutional Neural Networks Architecture: The distinctive architecture of a CNN is characterized by specialized layers that execute convolutional operations, pooling, and fully connected computations. These interconnected layers collaborate harmoniously to systematically process and abstract information from input data, resulting in remarkable performance across diverse computer vision tasks, encompassing image classification, object detection, and image segmentation. The architecture of CNN consists of three types of layers: Convolutional Layers, Pooling Layers, and Fully Connected Layers.

In a convolutional network, the convolutional layer takes precedence over the initial layer. Following the convolutional layers or pooling layers, the final layer is the fully connected layer. As the CNN progresses through these layers, its complexity increases, allowing for the identification of larger portions of the image. The initial layers primarily focus on detecting basic features such as colors and edges. As the image data passes

through the CNN's layers, it progressively discerns more extensive elements or shapes of the object until it ultimately recognizes the desired object [7].

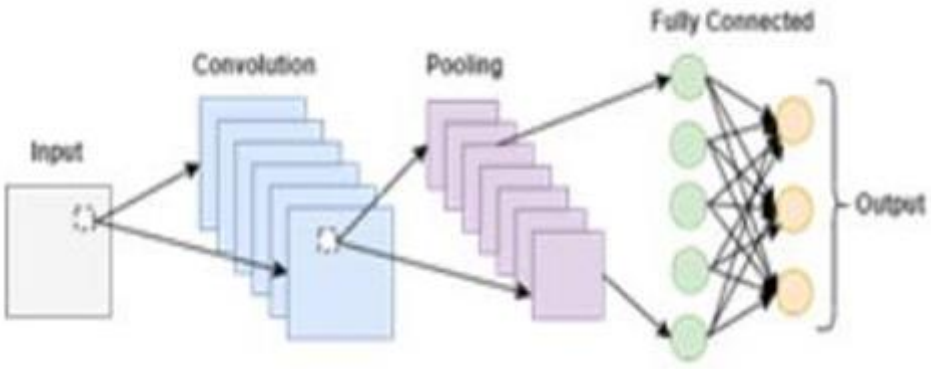


Figure 2: A Simple Convolutional Neural Network

In the context of classification, the fully connected layer utilizes the extracted features from previous layers, including their diverse filters. While ReLU functions are commonly employed in convolutional and pooling layers, the fully connected layers typically utilize a SoftMax activation function to accurately classify inputs. This activation function assigns a probability value ranging from 0 to 1, facilitating the classification process.

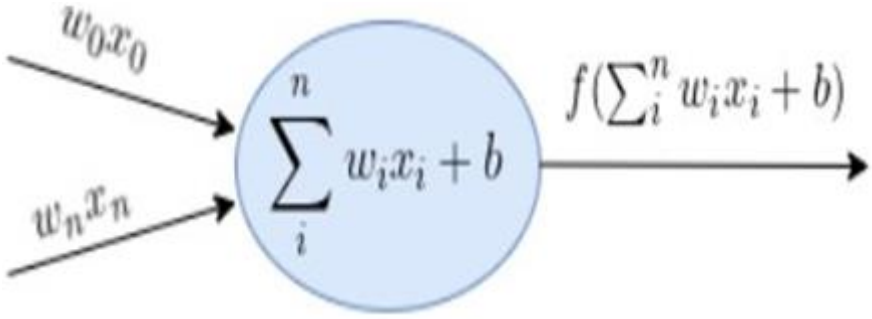


Fig 3. Mathematical Model of CNN

2) Convolutional Neural Networks (CNNs) in Self-Driving Cars: The ability to interpret the environment and make decisions based on that interpretation is a crucial component of a self-driving car system. Convolutional neural networks (CNNs) play a critical role in this aspect. These deep learning models are specifically designed for image recognition and analysis, making them well-suited for the context of a self-driving car.

By training a CNN to recognize and classify various objects and features in the environment, such as pedestrians, other vehicles, traffic signs, and road markings, the self-driving car can "see" its surroundings and make appropriate decisions. This includes adjusting its speed, changing lanes, and coming to a stop, among other actions [8][9].

B. AI Autopilot Cars and Simulators

AI autopilot cars and simulators play a crucial role in advancing and evaluating autonomous driving technology. These self-driving vehicles, with their enhanced safety and convenience, have gained substantial traction. Equipped with state-of-the-art sensors and software, they operate autonomously, utilizing machine learning algorithms to improve performance based on past experiences. Simulators serve as vital tools for developers, providing a secure and controlled environment to replicate diverse driving conditions and evaluate the

capabilities of self-driving cars. They enable the emulation of various scenarios, including different weather conditions, road layouts, and traffic situations. This allows developers to identify potential challenges and refine the performance of autonomous vehicles before deploying them on real-world roads. Moreover, simulators play a pivotal role in training artificial intelligence (AI) models that drive self-driving cars. By leveraging data from simulations, these AI models can effectively recognize and respond to a wide range of scenarios, reducing the reliance on real-world driving data and expediting the development process of self-driving cars.

1) Microsoft AirSim: AirSim, developed by Microsoft Research, is an advanced simulation platform tailored specifically for the development and testing of self-driving cars. It provides researchers and developers with a realistic and customizable environment to explore various aspects of autonomous driving. AirSim seamlessly integrates with Unreal Engine and Unity, delivering highly detailed and visually immersive virtual environments that faithfully replicate real-world scenarios.

One of the prominent features of AirSim is its ability to simulate sensor inputs, including cameras, lidars, and radars. This empowers developers to train and assess their self-driving algorithms using synthetic data, eliminating the need for costly and potentially hazardous real-world testing. By leveraging synthetic data, AirSim offers an accurate representation of the challenges and intricacies encountered on actual roads.

AirSim offers extensive control over the simulation environment, granting users the flexibility to customize diverse parameters such as weather conditions, traffic patterns, and road layouts. This adaptability enables developers to thoroughly test their algorithms in varied and demanding scenarios, ensuring robustness and adaptability in real-world situations.

Furthermore, AirSim provides a comprehensive set of APIs that facilitate seamless data collection, control, and analysis. Developers can effortlessly access and manipulate vehicle telemetry, sensor data, and ground truth information, enabling in-depth analysis and evaluation of their self-driving algorithms.

IV. REAL TIME STEERING ANGLE PREDICITON USING CNN

The research aims to develop a robust self-driving car model, necessitating a systematic approach to tackle inherent challenges. The front camera view served as the data, and the corresponding steering angle served as the label. A modular framework has been devised, comprising distinct modules with specific objectives and prerequisites. The modules include data collection and storage, data preprocessing, data augmentation, model training, and self-driving car execution. Each module plays a crucial role in the overall success of the work [10][11].

A. Data Collection and Storage

This critical module involves collecting expert demonstration data using the AirSim simulator and storing it for model training. The code initializes the AirSim client, captures car states and camera images upon a key press, and saves the data in subdirectories. This module forms the foundation for acquiring training data.

B. Data Preprocessing

Responsible for preparing expert demonstration data for model training, this module involves importing libraries, clearing directories, loading and standardizing state data, and resizing images. The pre-processed data is then saved in a designated directory, facilitating compatibility with the self-driving car model.

C. Data Augmentation

This module utilizes data augmentation to enhance the diversity of the training dataset. It creates right-turn patterns from left-turn patterns by applying mirror transformations and adjusting steering angles. The augmented data is saved in a new directory, contributing to the model's exposure to diverse and realistic training scenarios.

D. Model Training

Focused on training the self-driving car model, this module loads pre-processed data, defines a sequential model using Keras, compiles it with appropriate parameters, and trains it for ten epochs. The model's progress is displayed, and once trained, it is saved for future use during testing.

E. Self-Driving Car Control

This module utilizes the trained model to control the self-driving car in three modes: Full API Drive, Master Control Drive, and Refraining Model. In Full API Drive mode, the car is controlled entirely by the model's predictions. Master Control Drive mode allows the master to take over throttle and brake control, while Refraining Model mode refines the model by providing feedback and corrections.

F. System Architecture

The proposed self-driving car model utilizes a convolutional neural network (CNN) architecture. It consists of two convolutional layers with 3x3 filters and ReLU activation, followed by max pooling layers. The output is flattened, passed through a dropout layer, and then through two fully connected layers with ReLU activation. The final output layer uses linear activation for regression. The model is trained using mean squared error loss and the Adam optimizer. The architecture is implemented using the Keras Sequential API, with a summary available for visualization.

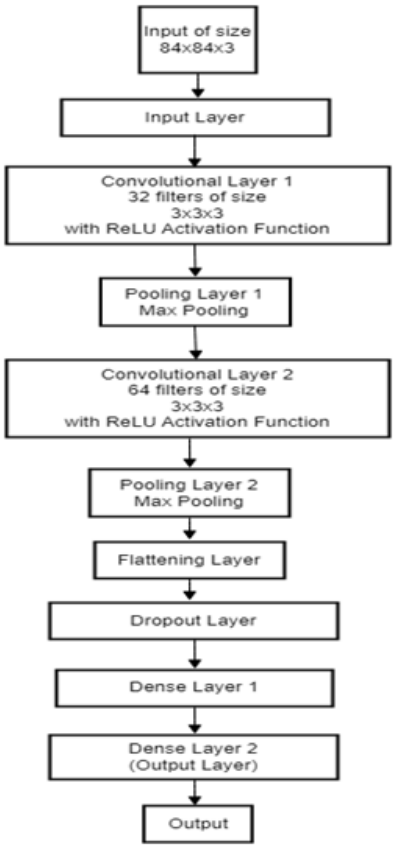


Figure 4: Architecture Diagram of the Model

V. RESULTS AND DISCUSSION

The obtained results indicate that the model demonstrated satisfactory performance on both the training and validation datasets, achieving accuracy scores of 0.8 and 0.93, respectively. However, a detailed analysis revealed that increasing the model's complexity did not yield any improvements. The phenomenon of overfitting was observed, where the model became excessively complex, relying on memorizing the training data rather than learning underlying patterns that generalize well to new data. This overfitting phenomenon can lead to excellent performance on the training data but subpar performance on unseen data. Therefore, finding the right balance between model complexity and performance is crucial for achieving optimal outcomes.

A. Key Conclusions Drawn From the Results Include

1. Increasing the model's complexity does not always enhance performance; a 2-layer model performed comparably to a 4-layer model.
2. Higher complexity can lead to overfitting, emphasizing the importance of model generalization.
3. Data augmentation proves to be beneficial, especially in scenarios where data is limited.
4. Performance metrics, in addition to accuracy, should be considered, as hardware limitations may impact the deployed model's effectiveness.

VI. CONCLUSION

Based on the obtained results, it can be concluded that the application of deep learning techniques proves effective in developing a self-driving car. The model, trained to predict the car's steering angle using input images, demonstrated good accuracy in both training and validation datasets. Notably, increasing the model's complexity did not significantly improve performance and led to overfitting. Thus, finding the right balance between model complexity and performance is essential for optimal results.

This research lays the groundwork for future enhancements, including expanding the model to predict other control parameters such as brake, speed, gear, and throttle. Real-world data collection through cameras and sensors on actual vehicles can further enhance model performance. The utilization of cloud-based environments for training and simulated scenarios for testing can expedite the development and deployment of these models in practical situations. Overall, the advancement of self-driving cars holds the potential to significantly improve road safety, alleviate traffic congestion, and offer a more comfortable and efficient mode of transportation.

VII. FUTURE ENHANCEMENT

The research, initially focused on predicting steering control, opens avenues for future enhancements by extending the model to predict multiple parameters simultaneously, such as braking, speed, gear, and throttle. This expansion can be achieved with a more powerful machine and an ample dataset. In real-world scenarios, the installation of cameras and sensors on vehicles can facilitate data collection from diverse sources. Preprocessing this data and utilizing a cloud-based environment for model training can enhance the model's capabilities. Simulated environment testing ensures accuracy before deploying models in actual vehicles.

Further improvements can be made by incorporating data from various driving scenarios and conditions, enhancing the model's accuracy and reliability. Additionally, exploring other deep learning techniques, such as reinforcement learning, holds potential for developing a more robust self-driving car system. The continuous exploration and integration of advanced methodologies contribute to the evolution and effectiveness of self-driving technologies.

These insights highlight the need for a nuanced approach to model development, where the trade-off between complexity and generalization is carefully considered to ensure optimal performance in real-world applications.

VIII. REFERENCES

[1] Jinila YB, Jabez J, Shyry SP. Virtual Self Driving Car using Improved Convolution Neural Networks. In2021 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT) 2021 Dec 10 (pp. 215-219). IEEE.

[2] Mallikarjuna M, Bhosle A. Self-Driving Car: Simulation of Highly Automated Vehicle Technology using Convolution Neural Networks. In2023 4th International Conference on Innovative Trends in Information Technology (ICITIIT) 2023 Feb 11 (pp. 1-5). IEEE.

[3] Mygapula DP, Adarsh S, VV SV, Soman KP. CNN based End to End Learning Steering Angle Prediction for Autonomous Electric Vehicle. In2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT) 2021 Sep 15 (pp. 1-6). IEEE.

[4] Yang J, Coughlin JF. In-vehicle technology for self-driving cars: Advantages and challenges for aging drivers. International Journal of Automotive Technology. 2014 Mar; 15:333-40.

[5] Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, Liu T, Wang X, Wang G, Cai J, Chen T. Recent advances in convolutional neural networks. Pattern recognition. 2018 May 1; 77:354-77.

[6] Babiker MA, Elawad MA, Ahmed AH. Convolutional neural network for a self-driving car in a virtual environment. In2019 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE) 2019 Sep 21 (pp. 1-6). IEEE.

[7] Indolia S, Goswami AK, Mishra SP, Asopa P. Conceptual understanding of convolutional neural network-a deep learning approach. Procedia computer science. 2018 Jan 1; 132:679-88.

[8] Babu Naik G, Ameta P, Baba Shayeer N, Rakesh B, Kavya Dravida S. Convolutional neural network based on self-driving autonomous vehicle (cnn). InInnovative Data Communication Technologies and Application: Proceedings of ICIDCA 2021 2022 Feb 24 (pp. 929-943). Singapore: Springer Nature Singapore.

[9] del Egio J, Bergasa LM, Romera E, Gómez Huélamo C, Araluce J, Barea R. Self-driving a car in simulation through a CNN. InAdvances in Physical Agents: Proceedings of the 19th International Workshop of Physical Agents (WAF 2018), November 22-23, 2018, Madrid, Spain 2019 (pp. 31-43). Springer International Publishing.

[10] Smolyakov MV, Frolov AI, Volkov VN, Stelmashchuk IV. Self-driving car steering angle prediction based on deep neural network an example of CarND udacity simulator. In2018 IEEE 12th international conference on application of information and communication technologies (AICT) 2018 Oct 17 (pp. 1-5). IEEE.

[11] Mohith G, Predicting Steering Angle for Self-Driving Vehicles. International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056, Volume: 09 Issue: 11 , Nov 2022, p-ISSN: 2395-0072.

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

X-ray Classification Using CNNs for Chest Disease Diagnosis

Dr. Ajesh F, Arunima S, Athul Sai, Aylin Mariam Johns, Bibisha Susan Mathew

Department of Artificial Intelligence and Machine Learning, Sree Buddha College of Engineering, Pattoor, Kerala, India

ABSTRACT

Medical imaging has been transformed by the advent of deep learning, more especially convolutional neural networks (CNN). It creates new avenues for imaging using chest X-rays. This study investigated the use of data from +detection and diagnosis of chest illnesses such pneumonia, TB, and COVID-19. Working with healthcare experts has an impact on treatment and adherence, despite obstacles such disparities in data and standards of interpretation. This research aims to close the gap between deep learning methods and practical clinical applications, potentially leading to better patient outcomes and diagnoses. A web-based picture loading and prediction system was also created to enhance the user experience for deep learning in neuroimaging.

Keywords: diagnosis, chest diagnosis, deep learning, convolutional neural network (CNN)

I. INTRODUCTION

X-ray imaging stands as a paramount method employed by radiologists to diagnose various bodily structures encompassing the heart, lungs, bones, blood vessels, and airways. This diagnostic tool proves highly effective in discerning a plethora of illnesses and abnormalities within the human body. Nevertheless, the World Health Organization (WHO) underscores the gravity of untreated chest ailments, which can lead to fatal consequences, contributing to millions of deaths globally.

The production of basic grayscale chest X-ray (CXR) images entails projecting X-rays onto the human body and imprinting the resultant image onto a steel plate. Despite the pivotal role played by radiologists in diagnosing thoracic diseases, the visual interpretation remains intricate and susceptible to errors. Research indicates that prolonged interpretation durations correlate with an elevated likelihood of erroneous diagnoses. Even seasoned radiologists may err, given the potential oversight of soft tissue and bone lesions along with associated symptoms.

The WHO underscores the imperative for precise and prompt treatment in combatting severe chest ailments responsible for claiming millions of lives annually. Among these, tuberculosis alone accounts for 1.4 million deaths each year, constituting a significant portion. Moreover, pneumonia emerges as a leading cause of mortality, claiming the lives of approximately 9 million children under the age of five worldwide. Notably, as of November 2022, the COVID-19 pandemic had resulted in over 6 million deaths globally, exacerbating the burden of chest-related illnesses.

In light to these challenges, there is an urgent need for innovative strategies to enhance the accuracy and efficacy of chest illness diagnosis. Deep learning methodologies, particularly convolutional neural networks (CNNs), leverage annotated CXR datasets to facilitate precise identification and categorization of various thoracic diseases. CNNs exhibit significant potential as tools for automating and augmenting the diagnostic process. With the aim to mitigate diagnostic errors, initiate treatment promptly, and ultimately improve patient outcomes, this study delves into the utilization of deep learning models for the accurate identification and classification of chest disorders.

II. MATERIALS AND METHODS

A. Related Works

The rise of deep learning (DL) presents a significant opportunity to streamline medical imaging diagnoses, potentially easing the workload for radiologists. Through the utilization of neural networks and deep learning methodologies, the accuracy of disease diagnoses in medical images has seen notable improvement, thanks to the availability of extensive datasets and robust computing resources. Several studies have explored the effectiveness of DL techniques in disease detection, leveraging diverse datasets like NIH, ChestX-ray14, VinDr-CXR, CheXpert, PLCO, and MIMIC-CXR. These investigations have highlighted the efficacy of DL-based computer-aided diagnosis (CAD) systems in identifying critical thoracic conditions.

Abiyev and Ma'aitah[*1*],for instance, delved into the use of deep convolutional neural networks (CNNs) for diagnosing chest diseases, demonstrating the ability of CNNs to distinguish between different chest ailments from X-ray images. Despite achieving impressive accuracy rates ranging from 80% to 95%, their study faced limitations due to the size of the dataset, potentially constraining the model's broader applicability.
Cha et al. *[2]* focused on assessing the effectiveness of a deep learning model in detecting operable lung cancer from chest radiographs, with an emphasis on early cancer identification. Although their study boasted a sensitivity of 85% and specificity of 90%, it primarily relied on data from a single institution, lacking diversity in its dataset.

Rasheed et al.*[3]*proposed a machine learning-based system for diagnosing COVID-19 from chest X-ray images, addressing the urgent need for effective diagnostic tools amid the COVID-19 pandemic. Despite achieving a high accuracy rate of 92% in discriminating COVID-19 cases from other respiratory illnesses, the study encountered challenges due to the scarcity of publicly available datasets for COVID-19 diagnosis, potentially introducing biases into the model.

Inbaraj et al. *[4]*described a novel machine-learning approach for segmenting and predicting tuberculosis using CXR images, aiming to automate the segmentation and detection of TB lesions. While reporting an 87% accuracy rate in tuberculosis prediction, their study was limited by a small dataset size and lacked external validation, which could impact the model's ability to generalize to broader contexts.

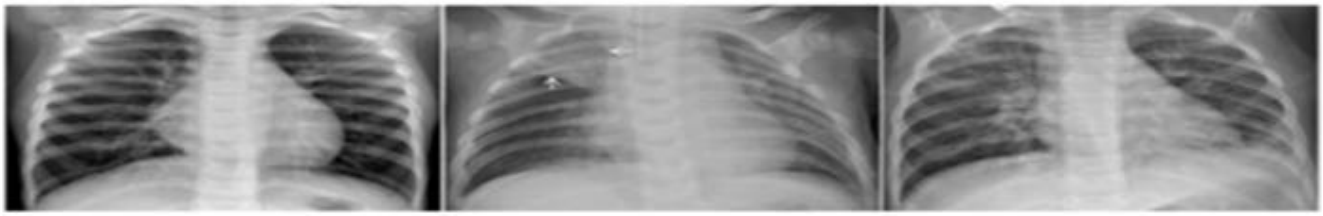
B. Disease diagnosed

1) COVID 19

Figure 1: X-ray images of (a) Normal, (b) (c) COVID-19

The virus responsible for COVID-19, officially named SARS-CoV-2, is a type of coronavirus knownto cause severe acute respiratory syndrome. Since the first reported case in late December 2019 in Wuhan, China, the disease has rapidly spread worldwide. The World Health Organization (WHO) officially designated the outbreak as a pandemic on March 11, 2020. Diagnosis of COVID-19 is primarily conducted using a method called reverse transcription polymerase chain reaction (RT-PCR). However, chest X-rays also play a crucial role in the early identification of COVID-19 due to their accessibility and ability to provide quick diagnostic images.

2) Pneumonia CXR Dataset

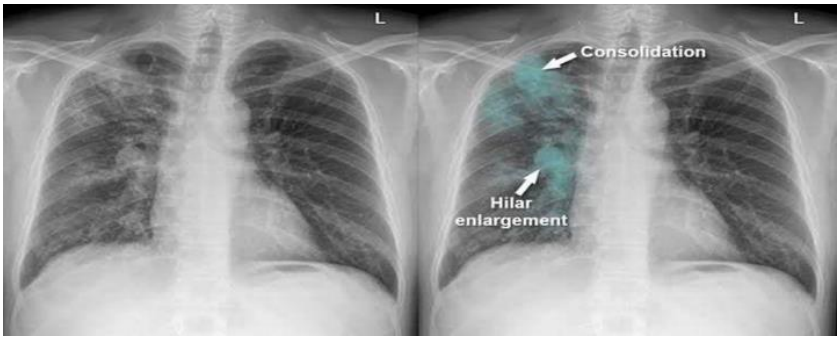


(a) (b) (c)

Figure 2: X-Ray images of (A) Normal, (B) Bacterial Pneumonia, (C) Viral Pneumonia

The dataset used in this investigation was sourced from the "Chest X-ray Images (pneumonia)" repository on Kaggle, curated by Paul Mooney. This dataset is instrumental for building models and conducting research on pneumonia identification, as it contains a plethora of chest X-ray images. These images are categorized into two main groups: those depicting pneumonia and those showing normal chest conditions. The dataset comprises images sourced from diverse origins, covering a wide spectrum of cases, including both paediatric and adult patients. This diverse range of cases provides a comprehensive understanding of how pneumonia presents across different age groups. Each image in the dataset has been meticulously labeled to distinguish between normal chest X-rays and those indicatingpneumonia.

3) Tuberculosis



(a) (b)

Figure. 3: X-Ray images of (A) Normal, (B) Tuberculosis

Mycobacterium tuberculosis remains the primary cause of tuberculosis (TB), a significant contributor to global mortality. Transmission occurs through airborne droplets, underscoring the importance of earlydetection and treatment to halt its spread. Chest X-ray (CXR) imaging plays a pivotal role in diagnosing TB by identifying characteristic lung abnormalities like infiltrates and nodules. However, confirmation often requires additional testing, such as sputum analysis. Treatment typically involves a multi-drug regimen administered over several months to ensure bacterial eradication.

TB control strategies encompass BCG vaccination, contact tracing, infection control in healthcare settings, and therapeutic interventions. Prioritizing early detection through CXR screening has the potential to mitigate TB's impact on health systems and communities worldwide, especially when coupled with prompt intervention and robust public health measures. Despite advancements, challenges like drug resistance and inadequate healthcare infrastructure persist, underscoring the ongoing need for global efforts in TB prevention, detection, and treatment.

Collaborative efforts in planning and resource allocation hold promise for significant progress in combating TB, ultimately saving lives, and mitigating the socioeconomic burden of the disease

C. Proposedsystem

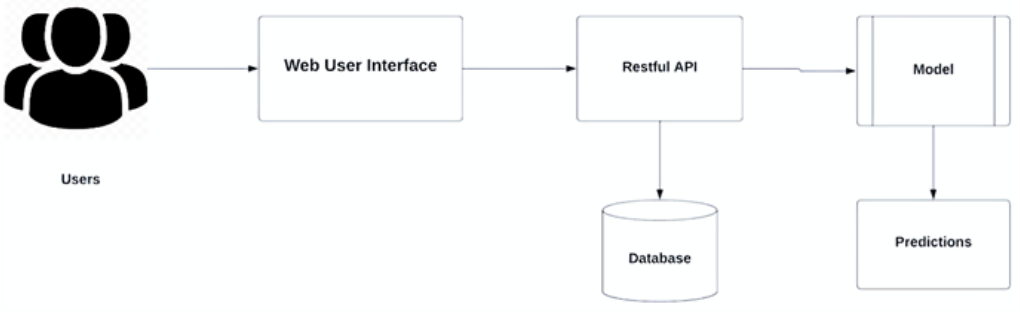


Figure 4: General Architecture

Chest X-rays play a crucial role in diagnosing diseases like tuberculosis and pneumonia. AI-powered diagnostic systems utilizing convolutional neural networks (CNNs) streamline interpretation processes and effectively detect subtle abnormalities. These CNN models are trained using diverse image datasets sourced from online repositories and hospital records. Preprocessing techniques are applied to enhance data quality, thereby optimizing CNN performance for accurate disease detection.

1) Training the CNN Model:

The CNN model undergoes a thorough training process using pre-processed chest X-ray images and the corresponding diagnosis of illnesses. A few essential actions are needed for this:

- Feature extraction: The CNN analyzes the images and extracts important features that recognize patterns and textures connected to disorders of the chest by applying convolutional filters.
- Feature Mapping: The CNN uses several pooling layers to transform the retrieved features into higher-level representations after feature extraction. By reducing complexity and maintaining important diagnostic data, this procedure optimizes the model.
- Classification: The CNN accurately classifies input images into many sickness categories using a fully linked layer and trained features.
- Backpropagation: By altering the model's weights in reaction to classification errors, this method maximizes the model's performance and increases projected accuracy.

The effectiveness of the trained CNN model is evaluated using an additional set of untrained chest X-ray images. Evaluation measures such as recall, accuracy, precision, and F1 score provide insight into the model's ability to distinguish between cases that are diseased and those that are not.

2) API Deployment:

The application programming interface (API) for the trained CNN model is used to expand the chest sickness detection system's usability. Through this user-friendly service, people may upload chest X-ray photos and receive disease forecasts promptly. The API is interoperable with Python, Java, and JavaScript, among other programming languages, guaranteeing wide accessibility and usage. Additionally, it makes integration with the current healthcare frameworks easier.

3) Future Directions:

With more advancements in AI, there are a plethora of potential to enhance the chest ailment detection system. Adding more and more datasets to the CNN model, along with utilizing state-of-the-art techniques like ensemble learning and transfer learning, can greatly increase both its accuracy and generalizability. Furthermore, a smooth integration of the system with Electronic Health Records (EHRs) can support a comprehensive approach to patient care. Furthermore, the development of mobile applications can facilitate the easy submission of X-rays via cell phones, thereby advancing early detection and timely treatment.

One excellent illustration of how AI could transform medical diagnostics is the chest ailment detection system. By automating and improving the processing of chest X-rays, this technology has the potential to revolutionise the way that sickness detection is approached and usher in a period of early diagnosis, improved patient outcomes, and cost-effective healthcare delivery.

4) Proposed method

The machine learning technique used in this process incorporates a number of crucial steps at various phases:

- Problem Definition: In this first stage, the work at hand is described in depth, and goals and guidelines are established for managing the machine learning research. It necessitates a thorough comprehension of the problem domain and the formulation of exact, measurable objectives to direct further effort.

- Data Collection and Preprocessing: After defining the issue, pertinent data is acquired from multiple sources to offer understanding of the result. Following collection, the data is carefully pre-processed to make sure it is suitable for machine learning analysis. This includes tasks like data cleansing and structuring.
- Model selection: The next step is to choose a design model that fits the goals of the study and the characteristics of the data using the previously collected data. At this point, factors like interpretability, scalability, and computational efficiency are carefully considered to evaluate whether the selected model is adequate for a certain task.
- Model evaluation and training: Following the selection of an appropriate model, an iterative process of training and evaluation is carried out. At this point, the model is exposed to data and learns about the connections and fundamental structure of the system. The model's performance is then rigorously evaluated using trustworthy data and metrics created for specific difficulties in order to ascertain the validity and generalizability of the model.
- Hyperparameter optimization: To improve the performance of the model, optimize the hyperparameters by examining the parameter space. Determine the optimal hyperparameter combinations using techniques like grid search, random search, or more complex optimization algorithms to maximize the predictive power of the model.
- Model deployment: After a successful installation, the training model is used to data in a real-world context to produce recommendations or support decision-making. This phase must be integrated with the current system and follow strict deployment rules in order to ensure dependability and capacity.
- Tracking and Monitoring: It's critical to closely monitor the model's performance after it's been implemented. Continuous evaluation facilitates prompt problem-solving and allows for gradual adaptation to shifting data dynamics or environmental conditions by identifying deviations or irregularities.

D. System specification

The software specs document outlines the system specifications, which encompass the system's purpose, user needs, and functional and non-functional requirements. By employing these techniques, we can ensure that the ultimate software system meets the requirements of its consumers.

1) Specifications for Development:

Table 1 contains the essential software requirements for the study's development.

TABLE 1. SPECIFICATIONS FOR DEVELOPMENT

Component	Required Specification
Operating System	WINDOWS 10 or above
Processor	Intel core i3 or above
Back end	Python CNN
IDE	Google Collab
RAM	4 GB

2) Specifications for Implementation:

The essential software specifications required for the implementation of the study are provided in table 2.

TABLE 2. SPECIFICATIONS FOR IMPLEMENTATION

Component	Required Specification
Operating System	Windows 10 or above
Processor	Intel core i3 or above
RAM	4 GB

E. Software Tools

Software development and office spaces provide an abundance of software tools, including libraries and programmes, to help complete tasks, simplify processes, and increase productivity. Each of these software tools was chosen to make a specific task easier, and they are all vital to the development process. This technology is essential for expediting tasks, optimizing processes, and ultimately guaranteeing the outcome.

1) Language and Libraries:

The languages and libraries used include the following:

- ***Python 3.10 :*** Launched on October 4, 2021, Python 3.10 is the stable version of the programming language that is readily available for download from the official website, python.org. A comprehensive Python code library has been built for the stress detection model. This selection of programming language is a reflection of Python's robust, effective, and extensive ecosystem of frameworks and libraries, all of which facilitate the implementation of intricate data processing and machine learning algorithms. finds relevant activities for the study of depression.
- ***Convolutional Neural Network*** *:* Convolutional Neural Networks (CNNs) are the pinnacle of deep learning architectures. They are meticulously designed to interpret structured grid data, especially images and videos, extremely well. Image analysis and pattern recognition have been revolutionized by CNNs, which are widely recognized for their exceptional capabilities in computer vision applications. CNNs are built on convolutional layers, which are crucial for feature extraction using filters. These layers aid in the network's ability to recognize complex structures in the input by gradually identifying intricate patterns. Next, pooling layers reduce spatial dimensions without sacrificing important features. Finally, fully coupled layers interpret and predict using learned properties.CNNs excel in autonomous feature learning and can create hierarchical representations from raw data, making them ideal for applications like object identification and image classification. Because of their shared weights and local connection, CNNs are great in translation-invariant pattern identification, which advances the science of computer vision and related fields.
- ***ResNet50 Model :*** We offer an enhanced convolutional neural network (CNN) model built on top of the ResNet50 architecture, which is widely used in picture categorization and other computer vision applications. Our augmentation involves adding custom dense layers to the ResNet50 base specifically to improve the model's performance for a given classification task. The two primary parts of the improved architecture are the additional custom dense layers and the fundamental ResNet50 model. To generate

the ResNet50 model, use the tf.keras.applications.The ImageNet dataset provides pre-trained weights for the initialization of the ResNet50 algorithm. This base model is a potent feature extractor that works with input photos that have dimensions of (224, 224, 3). It uses a series of convolutional and pooling layers in a cascade to capture intricate high-level features. Building on this foundation, we offer specialised deep layers made to manage the retrieved features for categorization. We make use of global average pooling to minimize spatial dimensions. To avoid overfitting, we then configured a dropout layer, ReLU activation, and a dense layer with 256 units. Finally, we create a probability distribution for the output classes using a softmax activated dense layer. The compilation with a personalized learning rate and categorical cross-entropy loss function is included in the model's optimization and assessment framework. With a learning rate of 0.0001, our model makes use of the Adam optimizer to dynamically adjust parameters in response to loss gradients. For multi-class classification tasks, categorical cross-entropy loss is a suitable choice since it accurately quantifies the dissimilarity between actual and anticipated class distributions. Accuracy metrics are used during training to examine the performance of the model by measuring the percentage of correctly identified samples. Our method preserves the strong feature extraction capabilities of the core ResNet50 architecture while enabling quick adaption to specific classification problems once the model architecture has been defined and training conditions have been set.

- ***Adam Optimizer :*** The Adam optimizer is widely recognized as a trailblazer in the domain of deep neural network training due to its innovative features that have propelled it to the forefront of optimization methodologies. Adam, called for its ability to adapt, seamlessly blends the greatest characteristics of AdaGrad and RMSprop to provide dynamic adaptability—a necessary component for adjusting learning rates during training. This adaptability is highly useful when navigating various gradients and achieving convergence under various criteria. Particularly impressive is Adam's grasp of momentum and integration speed strategies, which provide efficient problem-solving, raise convergence rates, and stabilize the training process.Because of its versatility and ease of integration, it has earned a solid reputation in the deep learning community as the go-to solution for optimizing neural network training, which eventually leads to the faster convergence and improved overall performance of the model.

2) Structure of the Proposed System:

To implement tiredness detection utilising facial recognition, Region of Interest (ROI), and Convolutional Neural Network (CNN), you can follow these general steps: Firstly, utilise a Data Acquisition Module to collect data from photos or videos that contain both sleep and alert faces. Optimally, the dataset should encompass a wide range of individuals and diverse scenarios, and we thoroughly cleanse and preregister the data. This stage encompasses various tasks, including adjusting the size of the image, standardising the lighting conditions, and employing methods like face detection and alignment to guarantee that the input aligns with the bite step. Utilise a facial recognition tool or the OpenCV library is used to detect and extract areas of interest (ROIs) that contain faces from each image or frame.

Annotation: Manually annotate regions of interest (ROIs) on the face as awake or asleep. It is imperative to perform this step in order to train the CNN model accurately. The datasets are categorised into training techniques, certificates, and examinations. The training process is employed to educate and instruct.

In the CNN model, the validation process is employed to adjust hyperparameters and assess the performance of the model, while the testing phase is utilised to measure the accuracy of the final model. Develop and train a Convolutional Neural Network (CNN) model utilising facial Regions of Interest (ROIs) from the training dataset.The model should be constructed using the ROI image output and the output prediction of sleep or wakefulness.

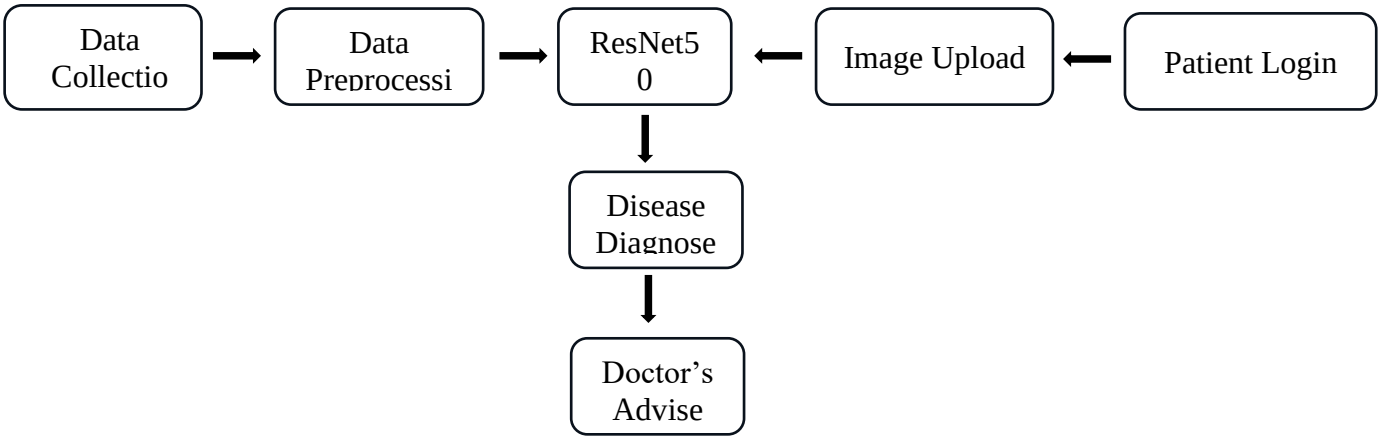


Figure 5: Architecture of Proposed System

3) Proposed Methodology:

In order to ensure datasets of superior quality and ethical integrity, the process of system development commences with meticulous data collection and preprocessing. This establishes the foundation for dependable creation, training, assessment, and incorporation of a collaborative doctor-patient interaction system to enhance patient care.

Data Collectionstage, a wide range of medical images are obtained to ensure comprehensive coverage of all relevant illnesses for the classification task. To maintain privacy and uphold ethical standards, it is crucial to strictly follow ethical procedures, such as gaining consent from patients or utilising datasets that have been curated in an ethical manner.

Data pre-processing involves the removal of unwanted or irrelevant photographs in order to optimise the quality of the dataset. In order to enhance the convergence and accuracy of the model, standardisation involves resizing photos to a uniform dimension, normalising pixel values, and dealing with noise and artefacts.

The model architecture involves the utilisation of custom dense layers specifically designed for the classification task. These layers are used to refine the process of feature extraction for medical image analysis. The ResNet50 basic model is then created and used for this purpose.

The training configuration involves using preloaded training data from ImageNet. The model is trained using Adam optimisation, categorical cross-entropy loss, and accuracy measurements. The layers of the ResNet50 model are modified as required while monitoring its performance.

Model evaluation indicators, such as recall, accuracy, precision, F1 score, and confusion matrix, provide vital insights into the performance of the illness identification model and aid in its practical development.

The Doctor Interaction System facilitates collaboration between medical practitioners and an automated model by providing a safe portal for examining disease descriptions, receiving comments, recommendations, and

prescriptions based on automated diagnoses. This promotes a collaborative and informed approach to patient care that prioritises accuracy and personalised treatments.

F. Implementation

The system is constructed using Convolutional Neural Networks (CNN). The enhanced ResNet50 model has bespoke dense layers that leverage the dense connectivity and feature reuse characteristics of ResNet50. These additional trainable layers are specifically designed to optimise performance for a particular classification task.

1) Module for authenticating users:

The user authentication model is responsible for safeguarding user identities, including those of patients and physicians. It serves as the foundation for managing security and integrity. The module functions as an entry point to medical platforms and ensures the security of sensitive data by carefully evaluating users to control their access.

The module utilises commonly accepted encryption algorithms and authentication processes to enhance this capacity. These measures not only surpass the most stringent criteria for safeguarding medical data, but they also enhance the security of users' personally identifiable information. The module offers a robust framework that ensures the privacy of patients while maintaining stringent control through the use of encryption technologies.

The module ensures user confidence and meets legal standards by prioritising security and safeguarding the privacy of medical records. Respect the confidentiality of medical records. The system ensures a secure and protected environment for the exchange of medical information, maintaining the utmost privacy and integrity through robust security measures.

2) Module for Uploading Patient Images:

The patient image upload module is essential for facilitating the secure submission of medical photographs by patients, hence aiding doctors in accurately assessing their ailments. The main goal is to provide a user-friendly platform that enables users to easily upload medical photographs. When these standards are fulfilled, accessibility and usability are guaranteed, and a variety of visual styles are utilised to enhance user comfort. The module has a rigorous verification process to meticulously assess the suitability and authenticity of provided photos. The verification technique ensures that only relevant and appropriate clinical images are processed, hence enhancing the overall accuracy of the sickness report model. This module is equipped with a variety of presets for editing and modelling uploaded photographs. This model has the capability to enhance subsequent training models and ensure uniformity in the dataset by performing the initial step. The patient image upload module has been meticulously designed to prioritise user experience, while ensuring the preservation of the quality and integrity of the patient's medical data. This eventually aids in accurate diagnosis.

3) Module for Identifying Diseases:

The ResNet50 standard was specifically created to accurately and precisely identify viruses by employing precise criteria. The main function of this model is to utilise deep learning algorithms to analyse medical images uploaded by patients. Furthermore, the implementation of the strategy requires utilising the dense network

model incorporated in the fundamental ResNet50 concept, in addition to merging the dense process. This integration enhances the level of detail, hence facilitating the analysis of medical imaging.

Upon extraction, the model is automatically identified as a predictive model. These justifications facilitate the interpretation and comprehension of the data, providing a more lucid understanding of the discovered therapies. Moreover, a seamless connection with physician interaction modules has been developed to assist doctors in diagnosing patients and exchanging information.

This integration facilitates collaborative work by providing clinicians with precise disease criteria that can be acted upon. This enables expedient provision of advice, recommendations, or medication delivery, hence promoting efficient and well-informed medical consultation. This integration bridges the knowledge disparity between interpersonal skills and knowledge of illnesses, so benefiting both patients and healthcare practitioners.

4) Medical Professional Communication Platform:

The module contains a variety of resources, including user manuals, training materials, and specific guidelines for medical professionals and patients. Users are given explicit instructions and guidance to encourage cautious utilisation, hence enhancing communication and familiarity with the system. Models play a vital role in empowering users, improving their capabilities, and encouraging adherence to best practices in the use of medical technology by prioritising the development of comprehensive knowledge.

The main role of this model is to prepare the process for referral, ensuring that its integration does not disrupt ongoing therapy. This involves comprehensive testing to identify and resolve potential issues and ensure the performance and reliability of the system. This technique ensures the security and integrity of medical records while also ensuring compliance with ethical norms, legal requirements, and healthcare standards.

The implementation of rigorous design principles throughout the distribution process ensures that the system is secure, fair, and in accordance with health protection guidelines. Deployment modules play a vital role in connecting development with the effective implementation of technology in real healthcare environments, especially when best practices and procedures are adhered to.

G. System testing

System testing is a crucial step in the entire testing process since it verifies the performance and functionality of a fully functional software product. This testing phase's main goal is to evaluate the system specifications in their entirety. All the components that have passed integration testing must be used as input in order to conduct system testing efficiently. Table 3 provides an outline of test scenarios to help direct the testing procedure.

TABLE 3. TEST CASES

Validation	Input	Response
Patient Login	User credentials	Successful
Image Upload	Picture	Successful
Disease Diagnose	ResNet50	Successful
Doctor Login	Doctor credential	Successful
Doctor Advise	Disease Diagnose	Successful

III. RESULTS AND DISCUSSIONS

In terms of application and evaluation, medical procedures have been warmly received, especially when it comes to diagnosing conditions and offering treatment alternatives. The sickness recognition system uses the improved ResNet50 model with the density approach, and it performs well in accurately identifying a range of disorders on patient-uploaded images. Based on excellent precision, recall, and F1 scores, the model demonstrated its capacity to distinguish between different diseases in several well-conducted trials.

Physician engagement portals are also essential for diagnosing patients and facilitating continuous doctor-to-doctor communication. Doctors may quickly and easily access patient data through the portal, including automatically produced disease descriptions, which helps with timely and informed clinical consultations. Process integration paves the door for improved patient care, as evidenced by electronic models and two-way doctor contact. This guarantees that recommendations for diagnosis and treatment are based on information and common sense.

Distribution modules are also essential to ensuring that the system is successfully implemented in the healthcare environment. The deployment process is faultless due to the careful alignment with existing systems, stringent testing procedures, and adherence to defined rules and controls. The system's comprehensive documentation, user-friendly interface, and rich educational materials all help to improve user experience and advance the connection between technical outcomes and applications in global health. Overall, it has been demonstrated via the application and evaluation of the therapeutic process that improving illness identification and therapy discussions can improve patient and clinical outcomes.

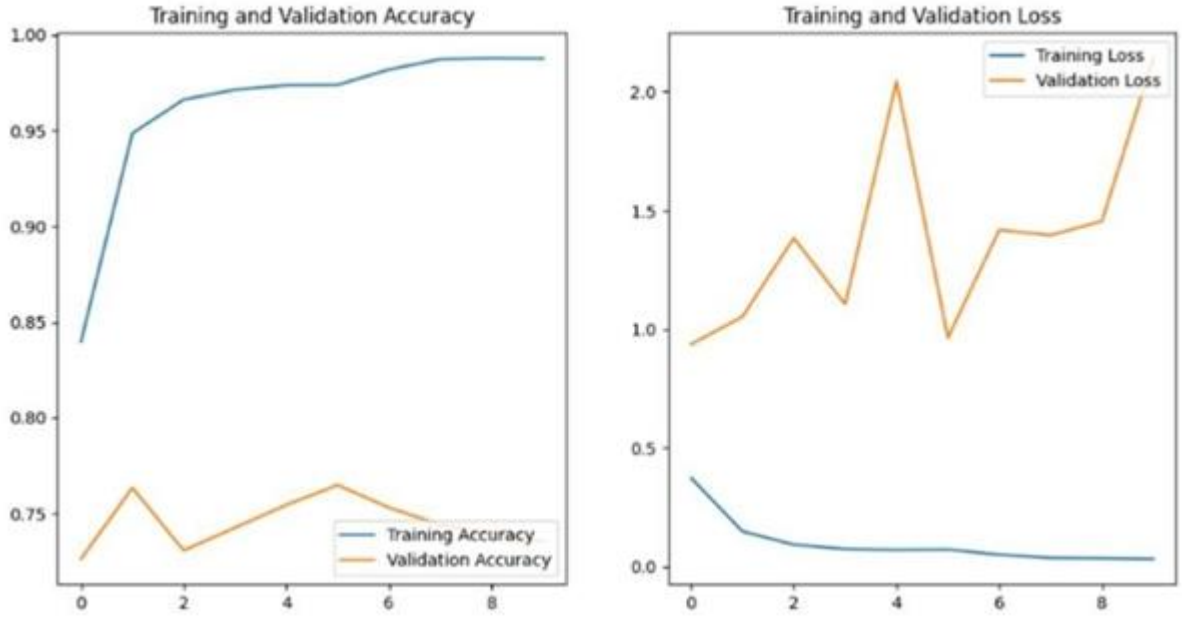
A. Comparison With Existing System

The proposed healthcare system is more advanced than the current practices, especially in terms of illness diagnosis and physician consultation. Unlike existing methods that heavily depend on human interpretation, the proposed solution integrates ResNet50 into a rigorous mechanism to generate reporting that is both unbiased and precise. This automation significantly enhances overall efficiency and accelerates the diagnosing process, while also simplifying the method. The new doctor interaction platform enables patients and clinicians to collaborate more efficiently and expeditiously. It offers a clear explanation of the ailment, allowing medical experts to promptly deliver advice, prescriptions, or suggestions to improve patient treatment.

Place security and compliance as the highest priority, and use strong mechanisms such as user authentication and encryption to ensure them. In order to maintain patient confidentiality, it is imperative to adhere rigorously to health protection regulations. The system's user-centric design, which incorporates an intuitive doctor interaction gateway and a patient-friendly image upload module, enhances the overall user experience due to its simplicity and convenience of use.

Moreover, the manual technique employed in the existing systems differs from the electronic identification system that replaces it. Ultimately, the medical system emerges as a viable and effective response to the challenges of modern healthcare, demonstrating exceptional performance in terms of precision, cooperation,

safeguarding, user satisfaction, and adaptability. The predicted revolutionary potential of this technology is expected to revolutionise healthcare by enhancing the safety and effectiveness of diagnostic and therapeutic consultations, leading to improved outcomes.



(a) (b)

Figure6:(a): Training and validation accuracy, (b):Training and Validation loss

B. Future Scope

It is imperative for this healthcare system to expand and enhance itself in order to stay abreast of developing technologies and evolving medical legislation. Firstly, by incorporating higher education institutions, such as state education or continuing education, the model can improve its transition to new medical knowledge while still maintaining strict patient confidentiality regulations. Further research and development could enhance the classification system to encompass a broad spectrum of treatments and uncommon medical conditions.

Moreover, utilise the Internet to broaden systems by integrating wearables, Internet of Things (IoT) devices, and real-time monitoring of patient health indicators. This will create opportunities for novel therapeutic methods and personalised treatment plans. Integrating with electronic health records (EHR) simplifies the sharing of information, promotes a comprehensive understanding of a patient's medical history, and facilitates informed decision-making by patients.

Moreover, promoting international collaboration can facilitate the endorsement of diverse models, streamline the process of constructing collective knowledge, and provide clinicians worldwide with a common platform. In order to achieve the best possible performance, it is required to continuously optimise the system, which involves discovering new models and fine-tuning the hyperparameters of the models.

The future of the healthcare business relies on a steadfast commitment to innovation, ethical decision-making, and collaboration. Ensure the continued relevance and effectiveness of these services as healthcare environments evolve.

IV. CONCLUSION

Overall, the work offers a new and efficient approach to accurately identifying and distinguishing lines by employing advanced deep learning technology, specifically ResNet50 with a dense layer. The study represents progress for a clinic that has traditionally lacked a doctor relations portal and a secure patient image submission module.

Instead than relying on traditional medical methods, automatic disease diagnosis is combined with prompt doctor-patient collaboration. Enhanced precision and productivity are achieved, while ensuring a sustainable and well-informed healthcare ecosystem. The synergy between collaboration and technology facilitates timely and knowledgeable therapeutic interventions, ultimately enhancing patient outcomes.

Moreover, strict adherence to the copy is essential to ensure the system's security, compliance with health regulations, protection of patient privacy, and adherence to corporate standards. The user-friendly interface guarantees overall satisfaction and active involvement in the system, continuously enhancing the user experience.

The model's inherent flexibility allows for continuous improvement, hence enhancing its precision and versatility across many clinical scenarios. The advantages of stability, accuracy, and coordination demonstrate how these elements might improve clinical consultation and sickness diagnosis.

This study serves as a vital foundation for demonstrating the potential and benefits of integrating state-of-the-art machine learning techniques into the field of medical care, as advancements in medical technology continue to progress. Successful completion and outstanding outcomes in patient care knowledge indicate a significant change, paving the way for additional breakthroughs at this critical intersection of medicine and technology.

V. REFERENCES

[1] R. Abiyev and M.K.S. Ma'aitah, "Deep Convolutional Neural Networks for Chest Diseases Detection," Journal of Healthcare Engineering, vol. 2018, article ID 4168538, 2018.

[2] M.J. Cha, M.J. Chung, J.H. Lee, and K.S. Lee, "Performance of deep learning model in detecting operable lung cancer with chest radiographs," Journal of Thoracic Imaging, vol. 34, pp. 86–91, 2019.

[3] J. Rasheed, A.A. Hameed, C. Djeddi, A. Jamil, and F. Al-Turjman, "A machine learning-based framework for diagnosis of COVID-19 from chest X-ray images," Interdisciplinary Sciences: Computational Life Sciences, vol. 13, pp. 103–117, 2021.

[4] X.A. Inbaraj, C. Villavicencio, J.J. Macrohon, J.H. Jeng, and J.G. Hsieh, "A Novel Machine Learning Approach for Tuberculosis Segmentation and Prediction Using Chest-X-ray (CXR) Images," Applied Sciences, vol. 11, article ID 9057, 2021.

[5] S. Santhi; M. Chairman Oral Disease Detection from Dental X-Ray Images using Densenet 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA).

[6] Shui-Hua Wang,Yu-Dong Zhang, DenseNet201-Based Deep Neural Network with Composite Learning Factor and Precomputation for Multiple Sclerosis Classification June 2020ACM Transactions on Multimedia Computing, Communications, and Applications.

[7] Ahmad Waleed Saleh, Gaurav Gupta, Surbhi B. Khan, "An Alzheimer's disease classification model using transfer learning Densenet with embedded healthcare decision support system "Decision Analytics Journal Volume 9, December 2023, 100348.

[8] Khan, A. Sohail, U. Zahoora and A.S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks Artif", Intell. Rev, vol. 53, pp. 5455-5516, 2020 [5] C. Shorten and T.M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning", J. Big Data, vol. 6, pp. 60, 2019.

[9] Estevão S. Gedraite and Murielle Hadad, "Investigation on the effect of a Gaussian Blur in image filtering and segmentation", Proceedings ELMAR-2011, pp. 393-396, 2011. [7] J. Patterson and A. Gibson, Deep learning: A practitioner's approach, Sebastopol:OReilly Media, Inc, 2017.

[10] Shruti Jadon, "A survey of loss functions for semantic segmentation", 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), pp. 1-7, 2020.

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

Culinary Community and Custom Recipe Discovery

Andrew Jose[1], David D M[2], Joice Mathew[2], Joyel Sani[2], Tom Alex[2]
[1]Assistant Professor, Department of Computer Science and Engineering Viswajyothi College of Engineering and Technology, Vazhakulam Ernakulam, Kerala, India
[2]Department of Computer Science and Engineering Viswajyothi College of Engineering and Technology, Vazhakulam Ernakulam, Kerala, India

ABSTRACT
This paper introduces a recipe recommendation application leveraging machine learning, particularly YOLO for real-time object detection, to enhance user culinary experiences. The system tailors recipes based on individual preferences and dietary restrictions, continually refining recommendations through iterative learning. A key innovation is the accurate detection and quantification of calorie content, addressing contemporary health concerns. The deployment of this application marks a transformative chapter, redefining the way users experience cooking with seamlessly integrated cutting-edge algorithms for personalized, health-conscious recipe recommendations.

Keywords: CNN, NLP, SVM

I. INTRODUCTION

The intersection of culinary arts and technology has witnessed a transformative shift with the advent of machine learning (ML) algorithms. This paper introduces a recipe recommendation application that pioneers a personalized culinary experience, anchored in ML methodologies. At its core, the application employs the You Only Look Once (YOLO) algorithm, renowned for its proficiency in real-time object detection and localization. This algorithm serves as the cornerstone for an innovative approach where users can simply capture ingredient images, and the system autonomously generates a curated list of recommended recipes.

The platform's commitment to user satisfaction is evident in its ability to decipher individual preferences, dietary restrictions, and evolving tastes. Through the meticulous analysis of user interactions and continuous refinement with vast datasets, the machine learning driven system ensures that each recipe recommendation aligns seamlessly with the user's unique culinary inclinations. This iterative learning process sets the stage for a highly tailored recipe curation process, redefining the user's culinary journey.

In response to the escalating concern over nutrition and diet-related health issues such as obesity and diabetes, the project aims to develop an innovative system capable of accurately detecting and quantifying the calorie content of food items. The significance of such a tool becomes paramount in empowering individuals to make informed dietary choices, contributing to a broader conversation about health-conscious living.

Upon deployment, the recipe recommendation application, empowered by machine learning, marks a pivotal moment in the culinary landscape. By seamlessly integrating cutting-edge algorithms, the platform aspires to redefine the way users experience cooking. Offering a world of tailored recipes and culinary exploration at users' fingertips, the application embodies the synergy between technological innovation and the artistry of cooking, promising a dynamic and personalized culinary journey.

II. LITERATURE SURVEY

A machine learning-based smart recipe recommendation system designed for a mobile edge computing (MEC) environment is developed. Traditional approaches often delegate such tasks to central cloud data centers, leading to increased response latency. The suggested system, however, offloads machine learning and recipe search tasks to a local MEC server, reducing both response latency and the computational burden on mobile user devices. It utilizes a chatbot-based interface and a Convolutional Neural Network (CNN) model for image recognition, aiming to enhance energy consumption efficiency and provide personalized recipe recommendations based on text and image inputs. The MEC server plays a pivotal role in handling computational tasks, ensuring privacy by keeping critical user information locally and interacting with external cloud servers only when necessary.

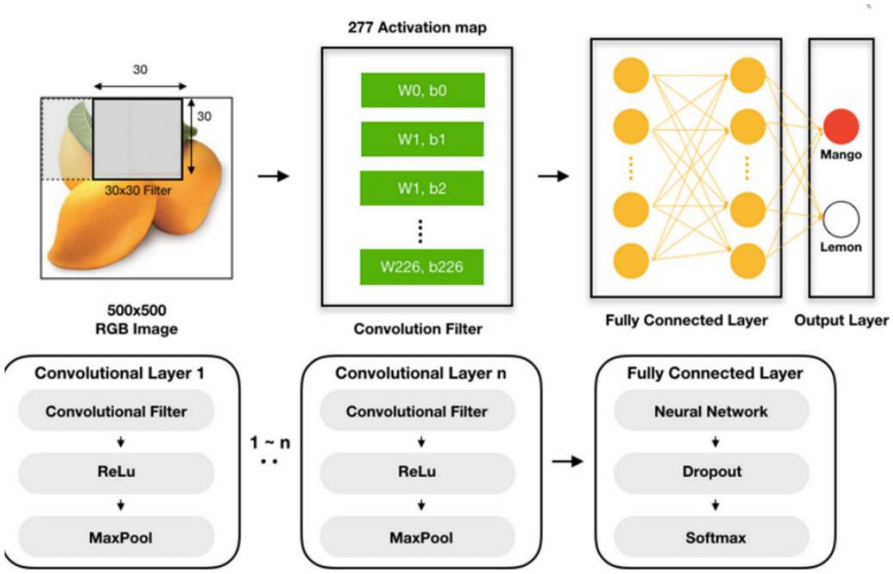


Figure 1: CNN model architecture

The MEC-enabled smart recipe recommendation system employs a CNN model for juice ingredient learning from user-submitted photographs. Fruit images, collected through web crawling, are resized and used as input data for training the CNN model. The architecture involves convolutional layers with ReLU activation, max pooling, and a fully connected layer for classification using the softmax function. The model is trained with a cross-entropy cost function through batch training. The system offers personalized juice recipe recommendations based on image recognition, promoting user engagement. Despite advantages such as MEC efficiency and privacy protection, there are challenges like data dependency, initial setup costs, model training

overheads, and user privacy concerns that need consideration in the implementation of this innovative recipe recommendation system [4].

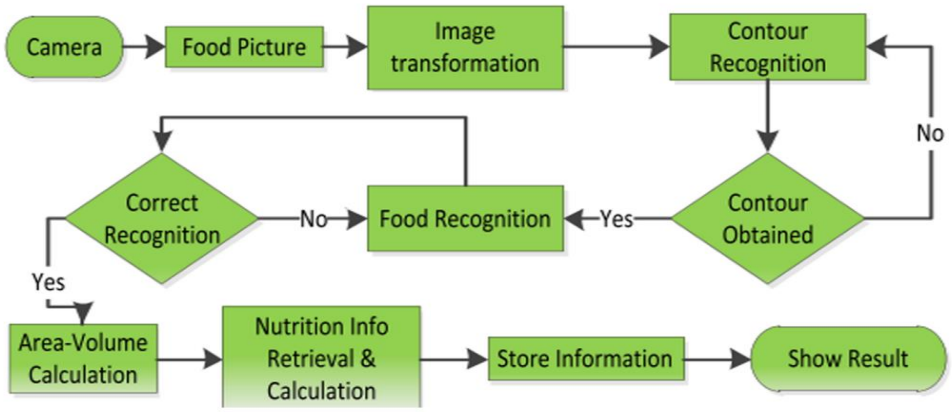
DeepNOVA, an advanced deep learning model, has been specifically crafted for the purpose of categorizing the nutritional quality of food items depicted in images. This task presents a significant challenge given the diverse and non-uniform nature of food images captured in various environments. The proposed methodology employs a comprehensive end-to-end deep learning framework, incorporating a tailored object detection model to identify and pinpoint the location of food items. The evaluation of the nutritional value for each identified item is carried out by classifying them into distinct NOVA groups, namely Unprocessed Food, Processed Culinary Ingredients, Processed Food, and Ultra-processed Food. The model undergoes training on a meticulously curated dataset, which encompasses both publicly available datasets and a custom dataset generated from images obtained through wearable cameras.

DeepNOVA's architecture involves the use of Darknet-53 to extract essential features from food images. The model's predictions focus on Bounding Boxes, Objectness Score, and Classifications, employing Anchor Boxes for better alignment. The paper emphasizes the importance of dataset preparation, including image pre-processing and data augmentation techniques. The experimental results showcase the performance of two models, SSD MobileNetV2 and EfficientDet D0, with a trade-off between processing speed and accuracy. Advantages of DeepNOVA include its innovative approach, end-to-end solution, and the use of diverse datasets. However, limitations include a relatively small training dataset, the labor-intensive nature of manual annotation, and potential ethnic/cultural bias limited to Tunisian food [7].

In the realm of calorie and nutrition measurement from food images, a novel system is proposed in this study. Employing image processing and nutritional fact tables, the system utilizes smartphone or tablet cameras to capture food images before and after consumption. The primary objective is to assist patients and dietitians in managing daily food intake, addressing concerns related to obesity and the necessity for precise food intake measurement. The study critically examines existing manual calorie measurement techniques, presenting the system's contributions, which involve addressing uncertainties in image-based calorie measurement, utilizing a diverse dataset, and incorporating multiple features for food recognition, including color, texture, size, and shape. The system's performance evaluation reveals commendable accuracy in area and volume measurement, as well as high precision in food recognition when all features are amalgamated. The study concludes by discussing potential applications and future enhancements of the proposed system [11].

Dataset preparation is a crucial phase in the system's development, involving the collection of an extensive dataset encompassing diverse conditions and food types. This comprehensive dataset ensures more meaningful and generalizable results, allowing the system to accurately recognize and measure various food portions. The preprocessing of images enhances their quality by eliminating noise or artifacts, facilitating effective image processing and segmentation. The meticulous data preparation phase plays a pivotal role in training and evaluating the system, enabling accurate measurement of calories and nutrition from food images. The subsequent sections delve into the utilization of machine learning models, particularly a support vector machine (SVM), for food classification. The SVM model, trained with feature vectors extracted during segmentation, contributes to accurate food identification, leading to precise measurement of calorie and nutrition content. The study also highlights the performance evaluation, emphasizing the system's accuracy through cross-validation and comparison with real values. Additionally, the architectural diagram, advantages,

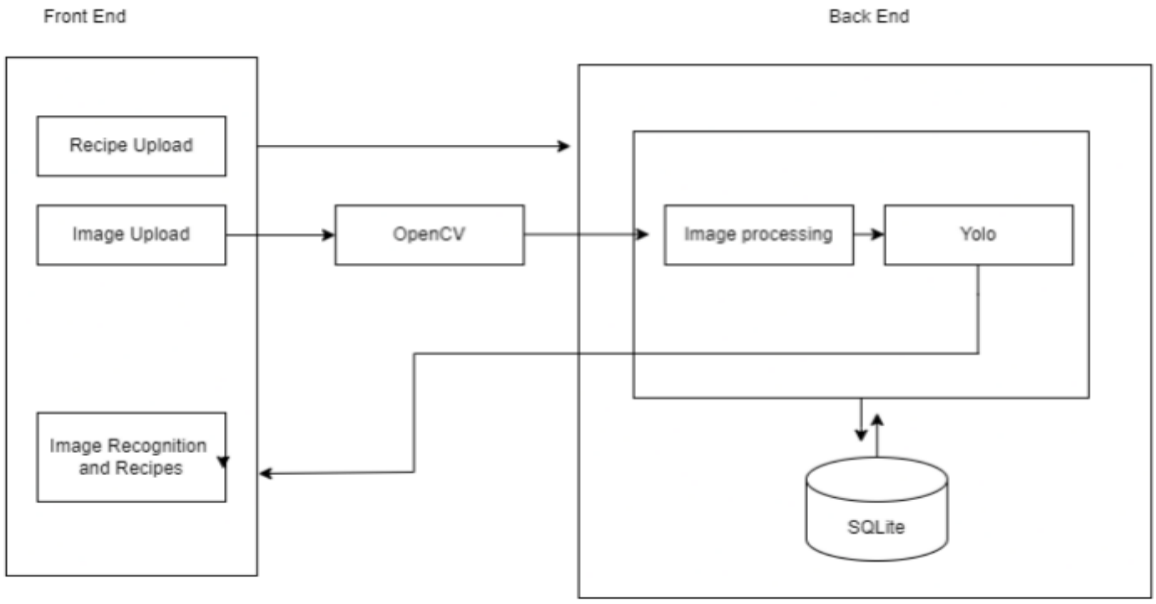
and disadvantages of the proposed system are presented, providing a comprehensive overview for literature survey purposes.



III. PROPOSED SYSTEM

The proposed system introduces Recipe in Social Media, a digital platform inspired by successful social media models, designed to cultivate a community of food enthusiasts and home cooks. Emphasizing engagement and information, the platform facilitates personalized profiles, recipe sharing, and community connections. Key objectives include addressing culinary pain points, such as dispersed recipe-sharing, generic recipes, and ingredient identification challenges. The proposed system offers functionalities like personalized profiles, recipe uploads with images, and community interaction, aided by advanced image recognition technology. Algorithms provide personalized recipe suggestions based on user preferences, creating a valuable resource with a vast database of recipes, cooking tips, and culinary knowledge for cooks of all skill levels.

1. Architecture



2. Implementation

The initiative represents a cutting-edge response to the contemporary demand for personalized culinary experiences. Utilizing machine learning as its foundational technology, the project combines tailored recipe suggestions with a dynamic online community platform for culinary enthusiasts. The machine learning component analyzes user interactions, incorporates feedback, and processes extensive datasets to continually enhance its recipe recommendations. This iterative learning process ensures that the system provides personalized recipe suggestions aligned with each user's unique culinary inclinations. The user-friendly approach allows individuals to simply photograph and upload images of ingredients, leaving the system to handle the rest, offering a variety of recommended dishes based on the uploaded images. The project's overarching goal is to create an innovative system capable of accurately detecting and quantifying the calorie content of food items, addressing the rising concerns over nutrition and diet-related health issues like obesity and diabetes.

The initiative's objectives encompass the development of an intelligent system capable of analyzing user-uploaded images of ingredients, offering personalized recipe recommendations considering individual preferences, dietary restrictions, and cooking habits. Simultaneously, the project aims to establish an online community where chefs can connect, share culinary creations, exchange ideas, and provide support to foster knowledge sharing and networking. The identified problems of dealing with varying image qualities, ensuring consistency in ingredient image recognition, addressing fragmented recipe experiences, and combating recipe overload underline the initiative's commitment to enhancing the user experience and streamlining the culinary exploration process across various platforms.

IV. COMPARATIVE STUDY

The below table shows the comparison between the different models reviewed for understanding the methodologies that could be used.

More table copy	***Advantages***	***Disadvantages***
Juice Recipe Recommendation System Using Machine Learning in MEC Environment.	Personalized Recommendations	Data Collection
DeepNOVA: Utilizing Deep Learning to Classify NOVA Categories in Food Images	It provides a comprehensive solution for food item detection and healthiness assessment	The model's applicability may be limited to Tunisian food
A Cooking Recommender for Ingredients and Recipes Using Set Transformer	provides unparalleled visibility and tracking capabilities.	Limited Dataset
Measuring Calorie and Nutrition From Food Image	Convenient food monitoring solution	Limited accuracy in measurement.

V. CONCLUSION

Recipe in Social Media is a specialized digital platform inspired by successful social media models, aiming to unite food enthusiasts and home cooks. It fosters an engaging community for recipe sharing, advanced vegetable recognition, personalized experiences, and provides cooking resources. The platform addresses issues in existing recipe-sharing platforms, such as dispersed content and generic recipes that overlook individual preferences and skill levels.

To achieve its goals, Recipe in Social Media allows users to create personalized profiles, share recipes with interactive features, and use advanced image recognition for vegetable identification. The platform's algorithms suggest recipes based on user preferences, catering to cooks of all skill levels and promoting creativity through user-generated content. This comprehensive approach aims to create a consolidated and interactive space for culinary enthusiasts to connect, share, and explore diverse recipes and cooking insights.

VI. REFERENCES

[1] SA ED S. ALAHMARI , (Member, IEEE), AND TAWFIQ SALEM. ”Food State Recognition Using Deep Learning.”in IEEE E Access , December 2022.

[2] Landu jiang 1,2, Bojia Qiu 2,Xue Liu 2 , Chenxi Huangi, AND Kunhui Lin . ”DeepFood: Food Image Analysis and Dietary Assessment via Deep Model.” in IEEE Access , February 2020.

[3] Yanai, Keiji, Takuma Maruyama, and Yoshiyuki Kawano. "A Cooking Recipe Recommendation System with Visual Recognition of Food Ingredients. " International Journal of Interactive Mobile Technologies 8.2 (2014).

[4] Mokdara, Tossawat, PriyakornPusawiro, and JaturonHarnsomburana. "Personalized food recommendation using deep neural network." 2018 Seventh ICT International Student Project Conference (ICT-ISPC). IEEE,2019

[5] Sachin, C., et al. "Vegetable Classification Using You Only Look Once Algorithm." 2019 International Conference on Cutting-edge Technologies in Engineering (ICon-CuTE). IEEE, 2019.

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

An Integrated Approach for Currency and Medicine Recognition for Blind People

Anju T[1], Aliya Ashraf[2], Ann Anna Aby[2], Devika S Dev[2], Rithu Babu[2]
[1]Assistant Professor Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Vazhakulam, Kerala, India
[2]Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Vazhakulam, Kerala, India

ABSTRACT
This project aims to empower visually impaired individuals by developing a dual-functionality application for currency and medicine detection. Using advanced technologies like Tesseract OCR and YOLO V3, the system ensures the reliable identification of Indian currency notes and medication details from images. The application converts extracted information into audio, facilitating independent decision-making for visually impaired users. This innovative solution addresses challenges related to fake currency detection and medication identification, fostering greater autonomy and security for individuals with visual impairments.
Keywords: YOLO V3, OCR, Text-to-Speech

I. INTRODUCTION

In a world dominated by visual information, individuals with visual impairments face numerous challenges in their daily lives. Simple tasks, such as identifying currency denominations or distinguishing between medication packages, can be daunting hurdles for the visually impaired. Recognizing the pressing need for inclusive solutions, our project aims to bridge this accessibility gap by developing a sophisticated Currency and Medicine Detection System tailored specifically for blind and visually impaired individuals.

The Currency and Medicine Detection System is designed to enhance the independence and confidence of the visually impaired community by providing real-time assistance in two critical areas: distinguishing different currency notes and identifying various medications. Leveraging cutting edge technology, our system combines computer vision, machine learning, and voice feedback to deliver a seamless user experience.

For the visually impaired, distinguishing between different currency notes can be a significant challenge, making transactions and financial independence more difficult. Our system employs advanced image recognition algorithms to accurately identify and announce the denomination of currency notes in real time. Users can simply point their smartphone or device equipped with a camera at the currency, and the system will provide instant audio feedback, enabling them to confidently manage their finances.

Managing medication is a critical aspect of healthcare, and visually impaired individuals often face obstacles in identifying and organizing their medications. Our project integrates a robust medicine detection feature that enables users to identify various medicines through their packaging. By capturing images of medicine packages using the devices camera, the system employs image recognition algorithms to analyse and announce the name and dosage of the medication. This empowers users to take the right medicine at the right time, promoting health and wellbeing.

This project not only addresses practical challenges faced by the visually impaired but also embodies the spirit of inclusivity and equal access to information. By leveraging technology to enhance Currency Detection for Visually Impaired People independence, our Currency and Medicine Detection System strives to contribute to a more inclusive and accessible world for all, regardless of visual abilities.

II. LITERATURE SURVEY

A. Scale-Invariant Feature Transform (SIFT) ALGORITHM

In the implementation of our project methodology, the initial step involved the creation of a database using the SQ Lite Database Management System to store both the names and images of medicines. SQ Lite, chosen for its embedded nature, eliminates the need for a client-server database engine. Currently, the database contains information for 10 medicines for testing, with flexibility for modification based on the user's prescription.

To identify the medicine, a pre-processing step is employed, where the input image is sharpened using the unsharp filter. This filter, implemented through the OpenCV library, subtracts an unsharp version of the image from the original, enhancing edges and widely used in photographic and printing applications. The primary objective of preprocessing is to eliminate unnecessary distortions and enhance critical image features for subsequent processing.

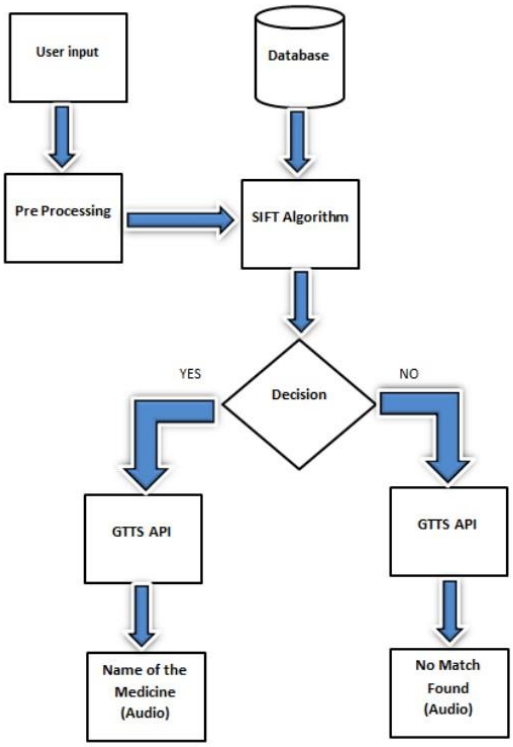
The SIFT algorithm is then applied to the pre-processed input image to identify key points. SIFT is known for its ability to detect and define local features. The attributes of the input image are compared against all images in the database using equations that describe the Difference of Gaussians (DoGs) and the convolution of the original image. Key points in SIFT are identified as maxima/minima of the DoGs.

If similarities are detected between the input image and those in the database, the system checks the number of good common points. If the count exceeds a predetermined threshold, indicating a substantial match, the name of the medicine is sent to Google's Text to Speech API. The API converts the medicine name into an audio signal, which is then relayed to a speaker. This approach allows blind individuals to independently identify the name of the medicine through auditory feedback, contributing to their autonomy in managing medication.

The process begins by inputting a tablet image into the preprocessing stage. The outcome of this preprocessing step is then directed to the Scale-Invariant Feature Transform (SIFT) Algorithm. Simultaneously, the data from the database is transmitted to the SIFT Algorithm. The SIFT Algorithm conducts a thorough examination by comparing the features of the input image with those stored in the database. Following this comparison, a decision is generated and sent to the decision maker.

If the decision is affirmative, indicating a successful match between the input and database images, the algorithm concludes that the identified medicine is correct. In such cases, the relevant data is sent to the Google Text-to-Speech (GTTS) API. The GTTS API then converts the medicine's name into an audio format, facilitating auditory recognition.

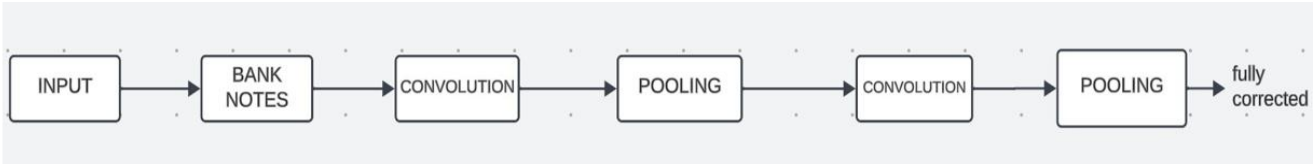
Conversely, if the comparison reveals no match between the input and database images, signifying an incorrect identification, the decision is categorized as wrong. This erroneous decision is then forwarded to the GTTS API, which produces an audio output to notify the user of the discrepancy. This sequential process ensures an effective and accurate mechanism for medicine identification, incorporating image preprocessing, feature analysis, and audio feedback to enhance accessibility and reliability.



B. CNN

VGG16 is a Convolutional Neural Network (CNN) model designed for extracting deep features from banknotes. Unlike traditional neural networks that use general matrix multiplication, CNNs focus on convolution operations. These operations involve sliding a function over another and measuring the integral of pointwise multiplication. The speed of this sliding window, known as strides, determines how quickly it moves, with a stride of two indicating movement by two pixels at a time. CNNs also employ pooling to down sample feature maps.

In the realm of deep neural networks, CNNs use fully connected layers for classification. The initial convolutional layer processes input in the form of a fixedsize range, typically a 224 x 224 RGB image. This image undergoes processing through a sequence of convolutional layers, where filters with a small receptive field (e.g., 3x3) are applied. Some configurations use 1x1 convolution filters for linear transformation of input channels, followed by non-linearity. The convolution stride is usually set to one pixel, preserving spatial resolution during convolution. Spatial pooling is achieved through multiple max-pooling layers after certain convolutional layers. Max-pooling involves down sampling over a 2x2 element window with a stride of 2. The network structure may include three fullyconnected (FC) layers, followed by an output softmax layer for comprehensive classification capabilities.



C. YOLO

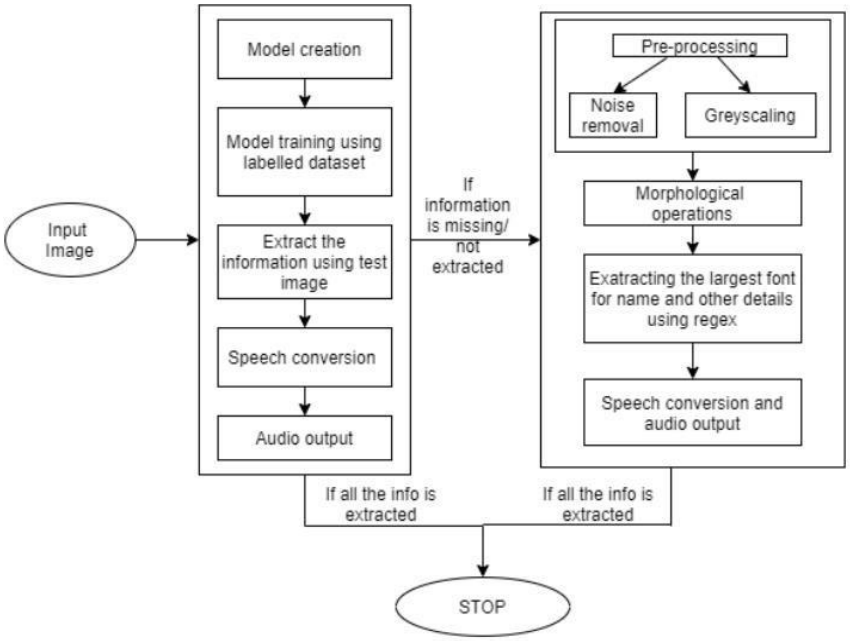
The project employs a model created through the Nanonets API, encompassing two primary building blocks: text detection and text recognition. The text detection component utilizes YOLO, a Single Shot Detector that predicts both the boundary box and class simultaneously. YOLO achieves real-time object detection but involves a trade-off between speed and accuracy compared to Regional-Based Detectors.

In the text detection process, data collection and augmentation are initiated, followed by annotation of images, where boxes are created, and each box is assigned its corresponding class. Training is conducted using darknet-53, and the iteration with the highest Mean Average Precision (mAP) score is selected for further data extraction.

For text recognition, Nanonets employs Tesseract 4 OCR. Once an image is passed through YOLO, it detects the necessary text regions and crops them from the image. Subsequently, these regions are passed to Tesseract for individual text extraction and storage of information.

In cases where the model fails to extract all required information, a step-by-step methodology is implemented. This involves image input, pre-processing (noise removal and grayscale operations), morphological operations (opening and closing), and text extraction using Tesseract OCR. The final output, obtained in dictionary format, includes details such as medicine name, dates, and MRP.

The step-by-step methodology continues with the final output being fed into the pyttsx3 (Python text-to-speech) library, converting the generated text into an audio output. The resulting audio output provides information in a structured format, including the name of the medicine and its associated details like MRP. This comprehensive approach ensures effective text extraction and conversion for individuals with visual impairments.



D. OCR

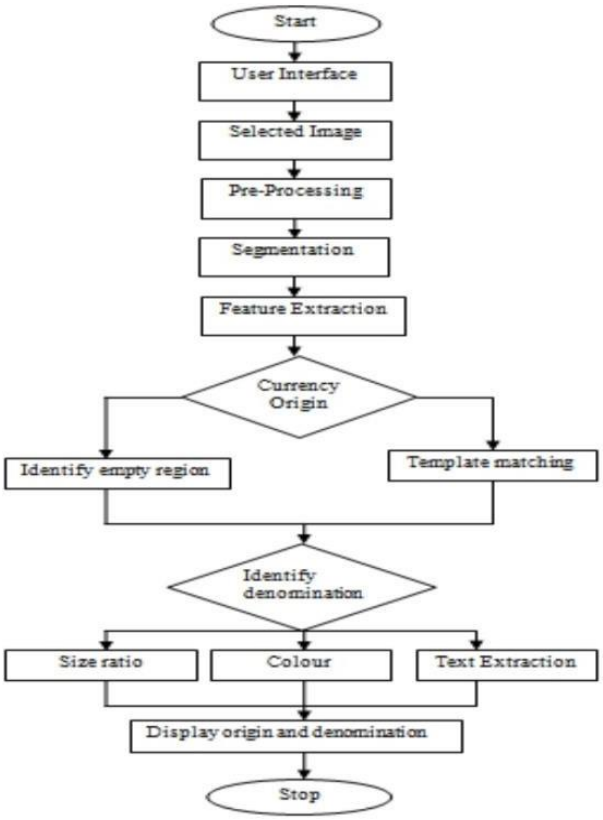
In the implementation of the Currency Recognition System, the process is structured into various sequential stages, starting with image retrieval. The initial stage involves acquiring images using methods like cameras or scanners, which retain crucial features such as size ratio, color, and text for subsequent processing.

Following image acquisition, the preprocessing step is crucial for enhancing the visual quality of the images and improving the impact on the dataset. This involves tasks like distortion correction, noise reduction, and conversion of the image to a binary format using adaptive thresholding. The primary objective is to prepare the images for further analysis and feature extraction.

After preprocessing, the system focuses on identifying empty regions on the banknote, categorizing them based on predefined areas and the ratio of black to white pixels in specific regions. This categorization helps group notes into left side empty, right side empty, center empty, or a separate category for notes with no significant space.

Moving forward, the system employs three approaches – size ratio, color, and text extraction – to identify the denomination of the note. Initially, it compares the size of the banknote with the known sizes of different denominations. If size alone is insufficient, the system uses k-means clustering on the banknote image to extract the dominant color, utilizing the LAB color space for accurate color perception. In cases where size and color methods fail, text extraction is applied to retrieve the denomination value typically written on the note.

The algorithm of the system is intricately designed in two parts, accommodating notes with and without significant empty regions. Upon identification and classification, the system matches input image features with available datasets to recognize the currency denomination. Subsequently, the recognized text is stored in script files, and a text-to-speech converter is utilized to generate an audio output. This audio output caters to visually impaired or blind users, allowing adjustments to preferences such as speech rate, volume, and language.



III. PROPOSED SYSTEM

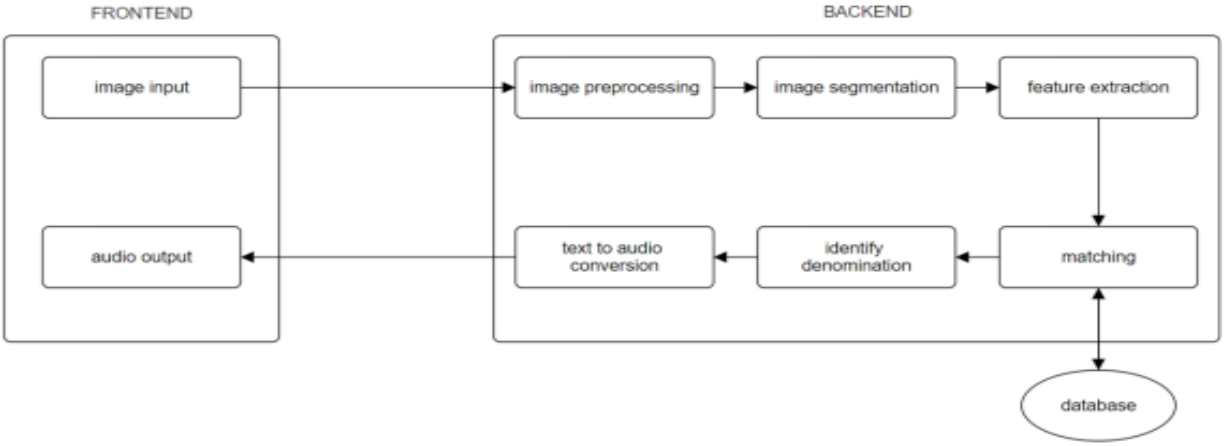
The proposed system Currency Detection for Visually Impaired People aims to empower visually impaired individuals by providing them with a reliable and userfriendly solution for both currency detection and medicine recognition.

The system employs advanced image processing algorithms to accurately identify and recognize various denominations of currency notes. Through the use of a mobile device or a dedicated camera, visually impaired users can effortlessly capture images of currency notes and receive real-time audio feedback regarding their value. Utilizing state of the art machine learning models, the system enables users to recognize and obtain information from medicine labels. By capturing images of medicine packaging or labels, users receive spoken details about the medication, dosage instructions, and other critical information, fostering independence in managing their health.

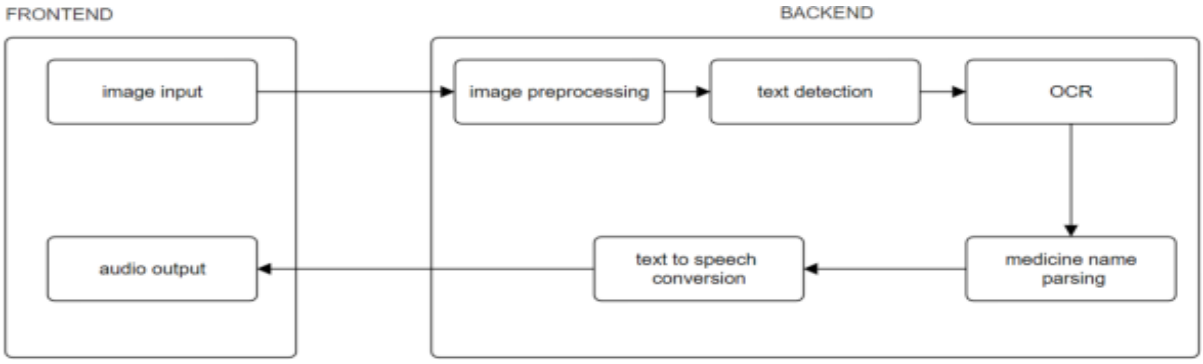
An integral part of the system is a robust text-to-speech engine that converts the recognized text into clear and understandable spoken output. This feature ensures that visually impaired users receive immediate and accurate information audibly, enhancing their ability to navigate the world around them independently. The system boasts an intuitive and accessible user interface, designed with the needs of visually impaired users in mind. Simple voice commands and tactile feedback options make the system easy to operate, promoting a seamless and enjoyable user experience.

A. Architecture

Architecturediagramforcurrencydetection:



Architecturediagramformedicinedetection:



B. Implementation

Forcurrencyrecognition:

- Image input: The user takes an image of the currency using a camera or uploads an imagefrom their device.
- Image preprocessing: The image is preprocessed to improve its qualityfor further processing. This mayinvolve:
- Noise reduction: To remove any noise that may interfere with the recognition process.
- Contrastenhancement:Tomaketheimageclearer and easier to analyze.
- Cropping and resizing: To focus on the region of interest (the currency) and ensurethe imageis the correct size for the CNN.
- Image segmentation: The process involves segmenting the image to separate the currency from its background, a task that can be accomplished using various methods and techniques.
- Colorthresholding:Basedontheknowncolor ranges of different currencies.
- Edgedetection:Toidentifytheedgesofthecurrency.
- Region-based segmentation: To group pixels that belong to the same object (the currency).
- Feature extraction: A CNN is used to extract features from the segmented image of the currency. TheCNNistrainedonalargedataset of images of different currencies and has learned to identify the features that are important for distinguishing between different currencies.These features may include:
- Shapes and patterns: Such as the unique patterns printed on each currency.
- Colors: The specific colors used in each denomination of a currency.
- Text:Thetextprintedonthecurrency, such asthe denomination or country of origin.
- Matching: The extracted features are compared to a database of features of known currencies. The system findstheclosest match inthedatabaseand identifies the currency accordingly.
- Denomination identification: If the currency is recognized successfully, the system can also identify the denomination based on the specific features extracted from the image.
- Text-to-audio conversion: (Optional) The denomination of the currencycan be converted to audio using a text-to-speech engine.

Output: The system outputs the recognized currency and its denomination in audio formats.

Formedicinerecognition:

- Image input: The user takes an image of the medicine using a camera or uploads an image from their device.
- Image preprocessing: The image may undergo preprocessing steps to improve the quality of the image and prepare it for further processing. This could involve noise reduction, contrast enhancement, or other techniques.
- Text detection: The system uses an object detection model to identify and locate any text regions within the image. This could be achieved through various methods, such as:

YOLO: A real-time object detection system that can identify objects in an image in real-time.

- SSD(SingleShotMultiBoxDetector):Another real-time object detection system that is efficient and accurate.
- EAST (Efficient and Accurate Scene Text Detection): A text detection model specifically designed for detecting text in natural scene images.
- OCR (Optical Character Recognition): Once the text regions are identified, an OCR engine is used toextract theactualtext content from those regions.
- Medicine name parsing: The extracted text is then parsed to identify the medicine name. This may involve using natural language processing (NLP) techniques to understand the context of the text and identify the relevant keywords.
- CNN (Convolutional Neural Network): The image is also analyzed by a CNN to extract visual features of the medicine. The CNN is trained on a large dataset of images of different medicines and has learned to identify features that are important for distinguishing between different medicines. These features may include:
- Shape and size: The overall shape and size ofthe medicine.
- Color: The color of the medicine and any markings or logos.
- Texture: The texture of the medicine, such as whether it is smooth, glossy, or bumpy.
- Information fusion: The information extracted from the text (medicine name) and the image (visual features) are fused to improve the accuracy of the recognition. This fusion can be done using various techniques, such as:
- Concatenation: Simply combining the text and image features into a single vector.
- Attention mechanism: This assigns different weights to different parts of the text and image features based on their importance for the task.
- Medicine recognition: Based on the fused information, the system determines the mostlikely medicine in the image.
- Text-to-speech conversion: (Optional) The name of the recognized medicine can be converted to speech using a text-to-speech engine.
- Output: The system outputs the recognized medicine name in audio formats.

IV. COMPARATIVE STUDY

Paper	Model	Advantages	Disadvantages
Recognition of Tablet using Blister Strip for Visually Impaired using SIFT Algorithm	SIFT Algorithm	Real-time Alerting	Dependency on Image Quality
Deep Learning-Based Indian Currency Detection for Visually Challenged using VGG16	CNN	Reduced dependency	Limited information
Medicine Identification Application for Visually Impaired People	YOLO	Adaptability	Complexity of Techniques
Implementation of Currency Recognition System	OCR	Improved accuracy	Only considers Indian notes

V. CONCLUSION

The development of a comprehensive application integrating both currency detection and medicine identification for blind individuals holds immense potential for enhancing their daily lives. By leveraging advanced computer vision algorithms and real-time data analysis, this innovative solution empowers visually impaired users to independently verify the authenticity of currency notes, safeguarding them from financial fraud, while also enabling them to accurately identify medications, ensuring their health and well-being. This dual-functionality app not only addresses critical safety and health concerns but also fosters greater autonomy and inclusivity for the blind community, marking a significant step toward a more equitable and secure future.

VI. REFERENCES

[1] R Shashidhar, V Sahana, Sudeshna Chakraborty, S B Puneeth, M Roopa" Recognition of Tablet using Blister Strip for Visually Impaired using SIFT Algorithm", Indian Journal of Science and Technology, https://indjst.org/articles/recognition-of-tablet-usingblister-strip-forvisually-impaired-using-sift-algorithm,2021

[2] Nijil Raj N, Anandu S Ram, Aneeta Binoo Joseph, Shabna S, "Deep Learning Indian Currency Detection for Visually Challenged using VGG16", International Journal of Recent Technology and Engineering(IJRTE),https://www.ijrte.org/wpcontent/uploads/paper s/v9i2/B3955079220.pdf,July 2020 Paper Model Advantages Disadvantages Recognition of Tablet using Blister Strip for Visually Impaired using SIFT Algorithm SIFT Algorithm Realtime Alerting Dependency on Image Quality Deep LearningBased Indian Currency Detection for Visually Challenged using VGG16 CNN Reduced dependency Limited information Medicine Identification Application for Visually Impaired People YOLO Adaptability Complexity of Techniques Implementation of Currency Recognition System OCR Improved accuracy Only considers Indian notes

[3] Angel Negi1, Aishwarya Bhure2, Dhanashree Patil3, Akshita Maskara4, Madhuri Bhalekar5, "Medicine Identification Application for Visually Impaired People", https://turcomat.org/index.php/turkbilmat/article/download/10346/7811/18432#:~:text=The%20proposed%20application%20is%20impl emented, to%20the%20server%20(TOMCAT),2021

[4] Arjun M. Shetty1, Shravya2, R. Manjesh3," Implementation of Currency Recognition System", International Journal of Research in Engineering, Science and Management Volume-3, Issue-5, May202

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

Bridging the Spectrum: AI and Virtual Assistants for Early Autism Detection

Ariya TK[1], Alvin Chackochan[2], Donjo Danty[2], Thomasukutty Benny[2]
[1]Assistant Professor, Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Ernakulam, Kerala, India
[2]Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Ernakulam, Kerala, India

ABSTRACT
Autism spectrum disorder (ASD) is a neurodevelopmental disorder associated with brain development that subsequently affects the physical appearance of the face. Early diagnosis and intervention are crucial for improving the quality of life for individuals with ASD. In recent years, machine learning and artificial intelligence (AI) techniques have shown great promise in aiding the early detection of ASD. This study presents an innovative approach that combines Convolutional Neural Networks (CNN) with a Virtual Assistant Tool for the prediction and detection of ASD. To ensure broader accessibility, the study has developed a web application using HTML, CSS, JavaScript, PHP, and Python. The core of the study lies in the amalgamation of Convolutional Neural Networks (CNNs) with a Virtual Assistant Tool. This dual approach harnesses the power of advanced image analysis through CNNs, while concurrently providing a user-friendly interface via the Virtual Assistant. This integration not only ensures robust diagnostic capabilities but also prioritizes user accessibility and experience. The integration of these technologies holds promise for early detection, enabling timely interventions and improved outcomes for individuals with ASD.

I. INTRODUCTION

Autism is a complex developmental condition that impacts various aspects of an individual's life and persists throughout their lifetime. Social impairments commonly associated with autism may involve challenges in maintaining eye contact, difficulty understanding and expressing emotions, and struggles with displaying appropriate facial expressions. Communication difficulties often observe as trouble initiating or sustaining conversations, a lack of responsiveness to one's name being called, and confusion with pronouns. Behavioural patterns may include unusual sensory responses such as heightened or reduced sensitivity to stimuli like smells, sounds, lights, textures, or touch, as well as fixation on activities or interests. Autism is recognized as a prevalent disorder with wide-reaching effects.

As per data from the National Informatics Centre, autism affects approximately 1 in 500 or more individuals in India. Studies conducted by researchers at the University of London have revealed that adults diagnosed with

autism spectrum disorder (ASD) often experience elevated levels of depression and face challenges in securing employment. Additionally, research indicates that ASD symptoms may exacerbate with age.

Most inquiries concerning autism spectrum disorder (ASD) rely on brain imaging datasets and recordings to differentiate individuals with autism. Presently, there is no definitive laboratory test for ASD. However, early detection of autism is crucial as it enables parents to implement strategies such as applied behaviour analysis therapy, thereby improving the child's quality of life. Researchers at the University of Missouri have identified distinct facial characteristics in children with autism. For example, children diagnosed with autism tend to have a broader upper face and a shorter middle region of the face compared to neurotypical children.

This paper introduces a method for identifying autism spectrum disorder (ASD) utilizing facial images of children. A comprehensive preprocessing approach is employed. Utilizing facial images for ASD classification proves to be more cost-effective compared to classification using brain imaging datasets. We conduct parallel classification utilizing MobileNet, InceptionV3, and InceptionResNetV2 models pre-trained on ImageNet.

II. LITERATURE SURVEY

A. 1-D CNN

In this paper [1] addresses different sorts of chronic and intense pathologies that influence the wellbeing of individuals. This proposed framework overwhelmingly centres on the advancement latest technology for the classification of eye, skin, tongue, and verbal abnormalities by Deep Learning which utilizes the transfer learning approach. The disease discovery and classification have been done by modifying the existing Convolutional neural network. The DenseNet121, MobileNetV2, RestNet152V2, are the CNN models for detecting eye, skin, and tongue illnesses and a 1-D CNN model has been utilized to identify voice irregularities.

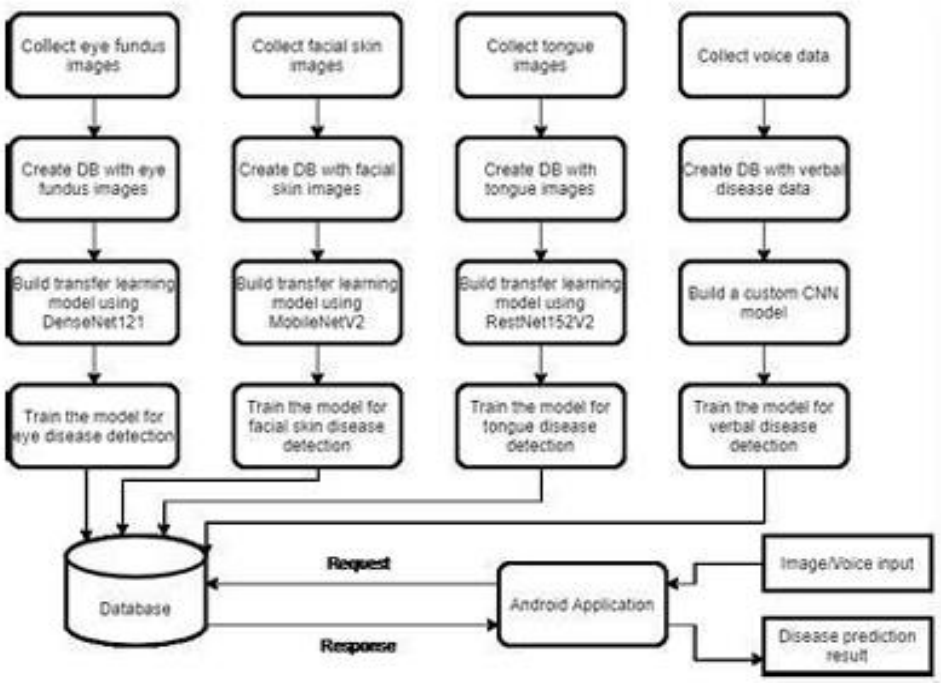


Fig. 1: Chronic Disease Detection System

The Initially the dataset was collected from the Kaggle dataset under the three main categories eye, skin, and tongue. The image preprocessing techniques such as Image segmentation, Image normalization, Image augmentation. These methods were adopted in order to enhance and optimize raw image data to improve the performance and accuracy of the model developed. The overall working of the system is shown in fig 1.

B. FACIAL ANALYSIS

In this proposed paper [2] we have a system for the classification of autism using various facial features. The main such features which has to be monitored for autistic individuals are expressions, action units, arousal, and valence.The neural basis of impairments in interpreting facial emotions in children and adolescents with autism spectrum disorders (ASD) describes the facial changes of autistic individuals. The faces of autistic individuals often exhibit distinct characteristics, including altered expressions, limited eye contact, and unique facial features, contributing to the diverse and individualized nature of autism spectrum disorder. Here the traditional convolutional neural network has been used for identifying these changes that arise in the autistic individuals. Here an end to end system has been developed for the detection of autistic individuals.

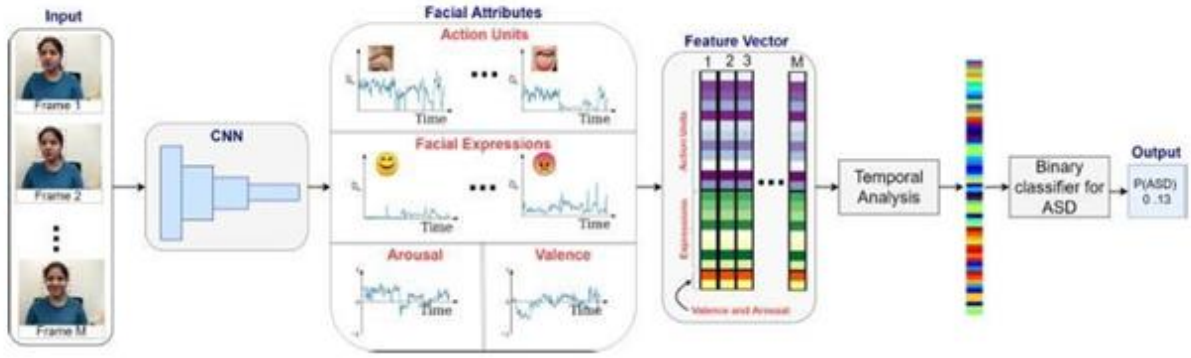


Fig. 2: ASD classification using facial attributes

In this , we utilize publicly available extensive datasets containing facial attributes crucial for effectively classifying natural images. This approach involves processing input videos, employing CNN model to extract four key facial features per each frame: facial expressions, action units , arousal, and valence on the user's detected face. The outcome is a comprehensive 22-dimensional feature vector, capturing probabilities associated with action units, expressions, and emotional attributes, yielding a personalized 9575 × 22-dimensional matrix for each participant. To enhance our analysis, temporal methods are employed, resulting in a more condensed 58- dimensional feature vector per participant. Despite the ex- tended execution time compared to other models in autism detection, our system achieves a statistically significant improvement of about 7% in ASD classification, boasting an F1 score of 76%. This was validated through thorough evaluation using the leave-one-out cross-validation on the limited dataset, demonstrating the system's effectiveness in identifying autism spectrum disorder. In Fig-2 the overall process that's involved in the classification of autistic individuals through facial features is provided.

C. TRANSFER LEARNING

This study [3] introduces an approach for autism detection utilizing facial images of children, aiming to identify this enduring developmental condition characterized by significant challenges in essential skills. Thus, models based on CNN works efficiently for the early detection of autistic individuals Here the classification relies on facial features such as a broader upper face and a shorter middle region compared to typically developing children, facilitating the identification of individuals with autism. Detection of ASD in which the input sample is a facial image is cheaper than the traditional ASD classification methodologies.

The dataset used in the paper is the Kaggle dataset. The dataset subdivided into autistic and non-autistic subgroups. MD5 and p-Hash algorithms have been used to clean and remove exact and near-exact duplicates of the training dataset collected. The algorithm operates much like the as illustrated The fig-3 provides the overall system architecture of the ASD detection system. The system is a type of binary classification which makes the prediction whether the user is affected with autism spectrum disorder.

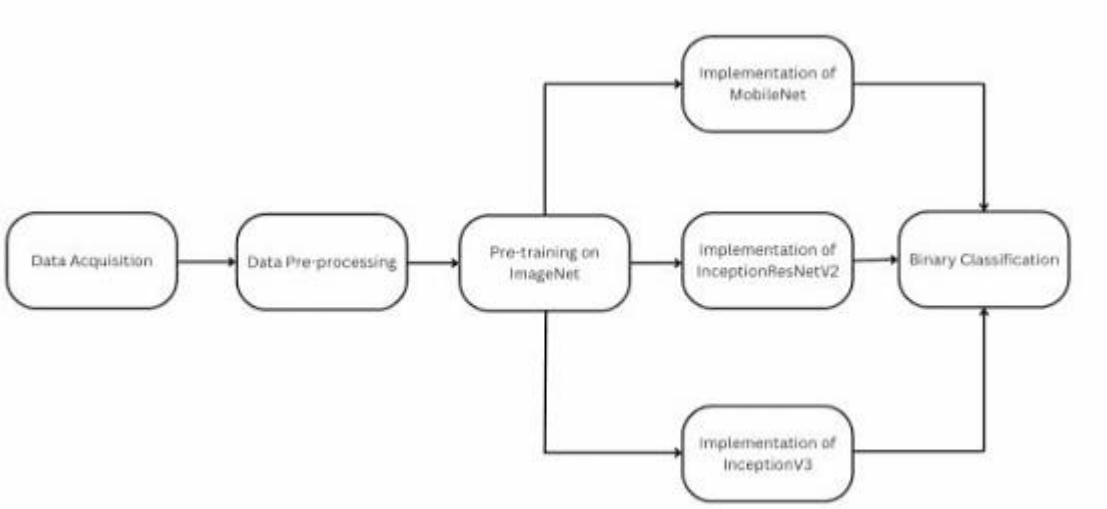


Fig. 3: Facial analysis system Workflow

The method involves pre-processing the dataset to remove duplicate images and then utilizing deep learning models for binary classification. Specifically, the study compares the performance of three different pre-trained models (Mobile Net, InceptionV3, and Inception Res NetV2) for the task of autism detection using facial images. Fig 3 depicts the working of the overall model.

To extract features, deep learning models are employed, as facial image features pertinent to autism are intricate to isolate, yet deep learning models excel in discerning subtle distinctions between the faces of autistic and non-autistic children. This will in turn help to develop an efficient system for the classification of ASD. The study emphasizes the need for a larger and varied dataset which in turn increases the model complexity.

D. TRANSFER LEARNING AND CONVOLUTIONAL NEURAL NETWORK

Here in this paper [4] a novel approach combining CNN and TL model has been designed for the rapid detection of Alzheimer's Disease. The pre- trained CNN model works on the pre-processed MRI images to extract the informative features for the prediction. The transfer learning model Alexnet is used for the efficient classification among the input given.

The region of interest i.e., the hippocampus of the brain is selected from the input and preprocessing of input is conducted. This is to ensure that the only essential part of the sample is processed for the classification. The CNN model comprises five convolutional layers with 3x3 filter sizes and 32 filters each, followed by a fully connected layer that produces the output.

Transfer Learning stands out as a widely recognized machine learning approach for image classification, enabling the application of previously gained knowledge from specific architectures to address fresh challenges . The transfer learning model AlexNet is a comprehensive network architecture featuring an input layer, five convolutional layers, and three fully connected layers. In Alzheimer's Disease Detection, we employ the AlexNet structure to classify images, where we adapt the network by replacing its original last layer, designed to classify 1000 categories, with a new layer tailored to the specific needs of our task, corresponding to the two classes of interest. This adjustment enables the model to learn features relevant to Alzheimer's disease classification from images sourced from the Oasis Dataset . The fig-4 specifies the overall structure of the Alexnet Transfer learning model. It specifies the general Alexnet model, and the changes adopted for the current classification. The findings indicate that the classification accuracy achieved with Transfer Learning (92.86%) surpasses that of the CNN approach (88.10%), underscoring the efficacy of Transfer Learning in transferring knowledge from a large dataset like AlexNet to a smaller one like Oasis. This transfer enhances model accuracy and reduces the learning duration needed.

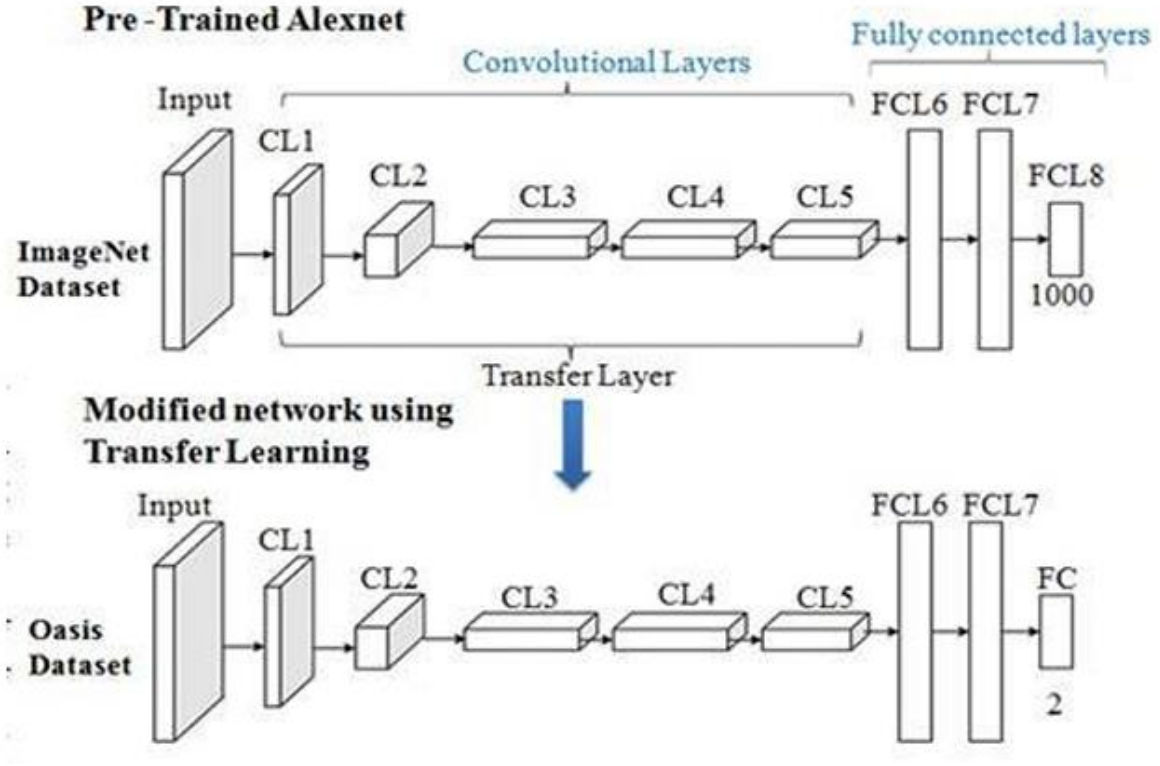


Fig. 4: Principle of Transfer Learning

III. PROPOSED SYSTEM

The proposed system for ASD detection and Virtual assistant using CNN is designed to detect autism from facial images and to propose an appropriate Virtual assistant support for the affected individuals. Beginning with the acquisition of diverse image datasets, this system employs rigorous preprocessing techniques to ensure the

clarity and fidelity of the collected data. Utilizing a meticulously trained Convolutional Neural Network (CNN) model, the system demonstrates its process in classifying these images into distinct categories, either autistic or non-autistic samples. This system, aims to have cost-effective, and accurate diagnoses, stands to revolutionize the field of autism spectrum disorder diagnosis, presenting a valuable resource for medical practitioners and researchers. The system includes various modules which add on the functionality to the system.

Upon logging into the system with the login credentials, users will be directed to the home page of the system, various functionalities have been provided in the home page. The user uploads a facial image of an individual to the system, The system preprocesses the image to ensure consistency and improve the accuracy of the detection model. This involves tasks such as resizing the image, cropping the face region, and normalizing the pixel values.

The system passes the extracted features to a machine learning model, which predicts the likelihood of the child having autism. The model may be a convolutional neural network (CNN), a support vector machine (SVM), or another type of machine learning algorithm. The system autism, as well as additional information about the features that the model identified as being relevant to the prediction. The admin of the page who controls all the action also has login credentials. After this, the admin monitors all the activities of the system. All the results generated by the user will be passed to the admin.

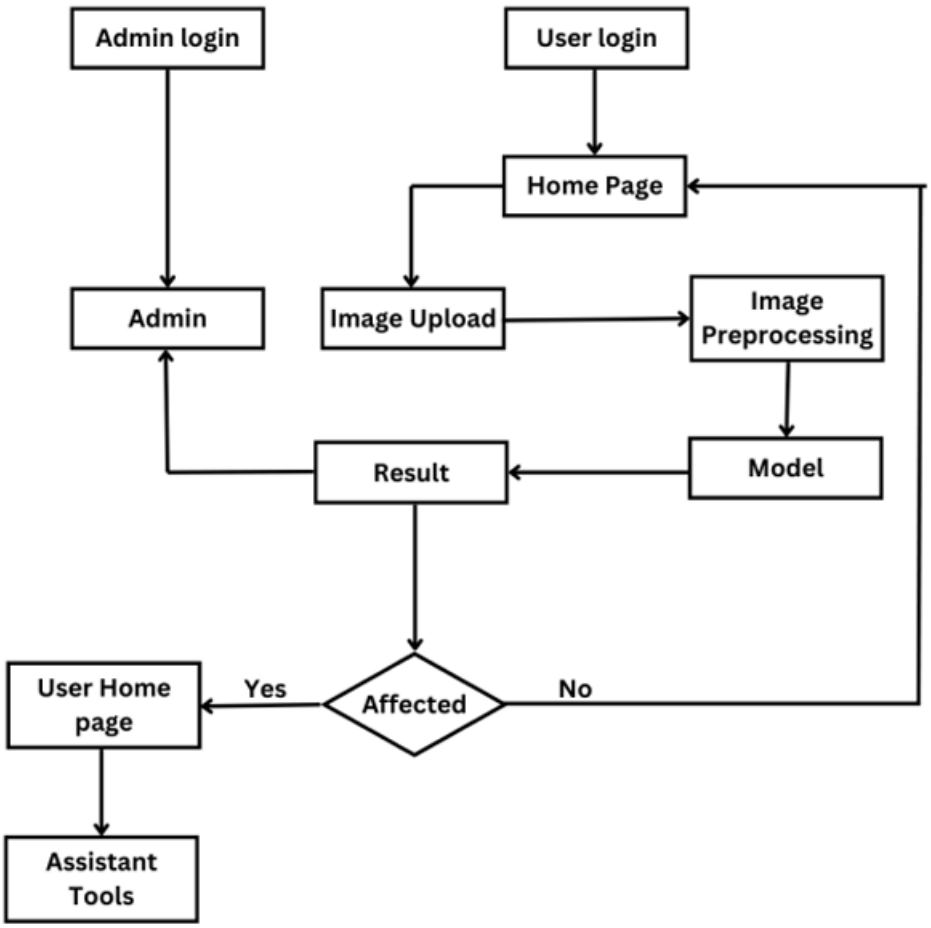


Fig. 5: Proposed System Architecture

After successfully classifying the autistic and non-autistic samples. The affected patients will be supported with a virtual assistant. The virtual assistant includes various task oriented modules which aim at the educational improvement of the autistic individuals. The proposed system architecture of the autism spectrum disorder and virtual assistant is depicted in fig-5. After the usage of the system the individual has to be capable of having a basic educational knowledge about alphabets and numbers. The virtual assistant also includes many tools for the overall betterment of the affected individual.

IV. COMPARATIVE STUDY

The below table shows the comparison between the different models reviewed for understanding the methodologies that could be used.

The table meticulously outlines the specific strengths and weaknesses inherent in each methodology, thereby assisting for understanding the efficient methodologies that could be used. By providing a comprehensive overview of the advantages and disadvantages associated with each method, it offers invaluable guidance in the meticulous selection of methodologies tailored to the research objectives and contextual requirements.

Authors	Method	Advantages	Disadvantages
Natasha Nigar, et al	Facial and verbal Disease Detection	Better performance than conventional CNN models without transfer learning.	Complexity in implementation
Sara Medhat, et al.	A facial expression assessment system for Autism spectrum disorder	Multi-task learning	Model Complexity
Noortaz Rezaona,et al.	Alzheimer's disease detection	Early Detection, CNN models can identify subtle biomarkers of Alzheimer's disease even in the early stages when symptoms may not be apparent.	Resource Intensive ,the training and fine tuning CNN models for Alzheimer's detection using MRI images require significant computational resources.
Christopher Ebuka Ojukwu	Detection Of autistic individuals	Better performance for the classification task which require high accuracy and low delay	LHigh computational intensity requires sophisticated algorithms, high dimensional datasets, and extensive model training.

Table 1 :Comparative Study

V. REFERENCES

[1] T D. M. Manage, A. M. I. S. Alahakoon, K. Weerathunga, T. Weeratunga,D. Lunugalage and H. De Silva, "Deep Transfer Learning Approach for Facial and Verbal Disease Detection,", 2021 .,

[2] Zewei Jiang, Shihong Yang,Zhenyang Liu,Yi Xu,Yujiang Xiong,Suting Qi ,Qingging Pang,Junzeng Xu,Fangping Liu and Tao X. “Coupling machine learning and weather forecast to predict farmland flood disaster: A case study in Yangtze River basin” in Elsevier, September 2022.

[3] N Y. Khosla, P. Ramachandra and N. Chaitra, ”Detection of autistic individuals using facial images and deep learning,”, 2021

[4] M. Zaabi, N. Smaoui, H. Derbel and W. Hariri, ”Alzheimer’s disease detection using convolutional neural networks and transfer learning based methods,”, 2020

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

An Approach on Real-time Audio to Sign Language Translator

Asha Joseph[1], Sona Xavier[2], Ashly Shaji[2], Miliya Elias[2], Shifin Vincent[5]
[1]Assistant Professor, Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Muvattupuzha, Kerala, India
[2]Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Muvattupuzha, Kerala, India

ABSTRACT
In an age characterized by technological advancements, the drive towards achieving inclusive communication takes a significant step forward with the introduction of a real-time audio-to-sign language translator. This innovative system aims to break down barriers for individuals with hearing impairments by establishing a dynamic link between spoken and sign languages. Communication stands as a fundamental cornerstone of human connection, posing considerable challenges for those with hearing impairments within traditional interaction modes. Acknowledging the necessity for a solution that surpasses these obstacles, the concept of a real-time audio-to-sign language translator emerges as a promising avenue to cultivate a more inclusive and accessible environment. This comprehensive system encompasses speech recognition, text generation, and output generation, providing a holistic approach to bridging the communication gap. The audio-to-sign language translator employs a combination of machine learning techniques to seamlessly generate the output, representing a sophisticated fusion of technology and linguistic accessibility. This innovative solution not only signifies a technological leap but also underscores a commitment to fostering an environment where communication is truly inclusive and barrier-free for individuals with hearing impairments.

Keywords: Speech Recognition, Machine Learning

I. INTRODUCTION

The real-time Audio to Sign Language Translator representsa groundbreaking leap in technological innovation, offeringa revolutionary solution to bridge the communication gapbetween the deaf and hard of hearing community and thosewho primarily communicate through spoken language. Thiscutting-edge system is built on advanced technologies likemachine learning, providing an instantaneous conversion ofspoken words into sign language. The seamless integrationof these technologies enables the translator to analyse andinterpret audio input in real time, generating live sign language animations that serve as a swift and accurate visualrepresentation of the spoken content.

One of the key strengths of this transformative technology lies in its ability to facilitate more inclusivecommunication. By offering a real-time translation ofspoken language into sign language, individuals

withhearing impairments can actively participate inconversations, educational settings, and social interactionswith unparalleled responsiveness. This not only enhancestheir overall communication experience but also empowersthem to engage more fully in various aspects of life thatmight have previously posed challenges.

The real-time Audio to Sign Language Translatorbecomes an invaluable tool, ensuring that individuals withhearing impairments have equal access to information andinstruction. The live sign language generated by the systemallows a dynamic and interactive learning experience,fostering a more inclusive environment. Moreover, in socialinteractions, this technology promotes a sense of connectionand understanding, breaking down barriers and fostering amore inclusive society.

Some of the key technologies integral to audio tosign language conversion include Neural MachineTranslation (NMT) mechanisms and computer visiontechnologies. NMT [5] seamlessly translates linguisticcontent into precise sign language gestures, effectivelybridging the gap between spoken language and visualexpression. Concurrently, computer vision [6] algorithmsinterpret visual cues derived from the audio source,augmenting the system's capacity to recognize and translatefacial expressions, gestures, and other visual componentsintegral to comprehensive sign language interpretation. Theharmonious interplay of NMT and computer vision servesto elevate the effectiveness and versatility of audio to signlanguage conversion systems.

II. LITERATURE SURVEY

A. CONVOLUTIONAL NEURAL NETWORK

Communication is an inherent right for all individuals, yet those with hearing impairments face a considerable gap in a world dominated by spoken language. The ability to express thoughts, emotions, and ideas through speech is often taken for granted, leaving those dependent on sign language at a distinct disadvantage in various aspects of daily life. Recognizing this disparity, there is a crucial need to develop a sign language conversion system to enhance communication.

[1]The CNN comprises three primary layers: the Convolutional layer, the Pooling layer, and the output layer. Additional layers, such as an Activation function and a Dropout layer, contribute to the classification process.The input audio is fed into the first convolution layer, generating an activation map through the extraction of relevant features by the layer's filters. Each filter provides distinct features crucial for accurate class prediction.

To preserve the image size, either padding or valid padding is employed in the convolution layer, effectively reducing the number of features. Pooling layers are subsequently added to further diminish parameters. Multiple convolution and pooling layers precede the final prediction. Convolutional layers play a pivotal role in extracting essential features, and the output layer, a fully connected layer, transforms input from preceding layers to produce the final output.

Ultimately, the speech signal input undergoes mapping to the corresponding sign language letters, facilitating effective communication for individuals with hearing impairments.

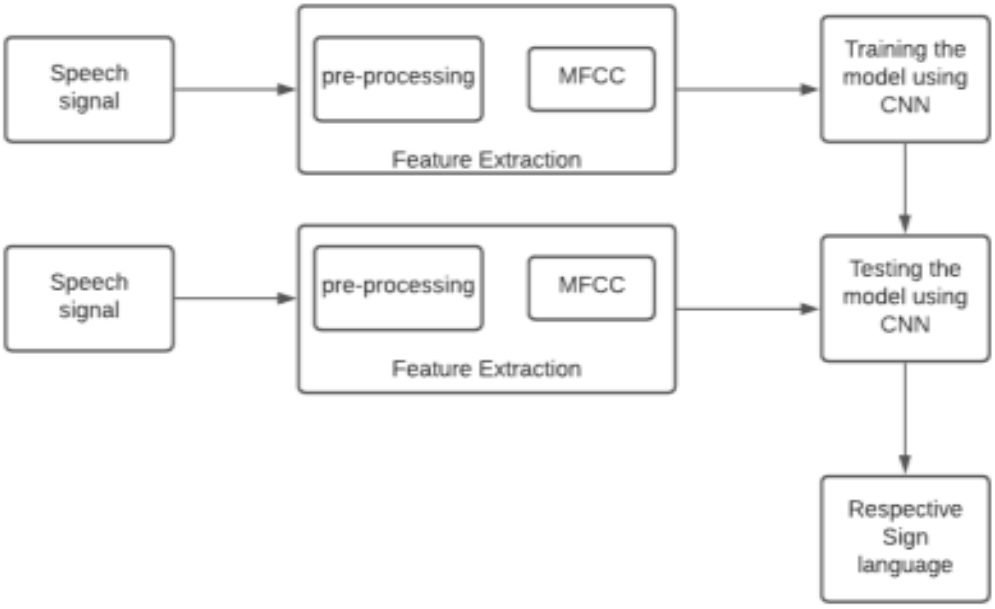


Figure 1: The architecture of speech to sign language conversion

B. NATURAL LANGUAGE PROCESSING

Effective human communication functions as a coordinated symphony, synchronizing the rhythm of communication tointerlace thoughts, emotions, and ideas. It is a fundamental cornerstone of connectivity with the world and those in close proximity. However, within this intricate fabric of human interaction, individuals grappling with hearing disabilities encounter substantial challenges. In this context, the Audio to Sign Language Converter emerges as an innovative system intricately crafted to dismantle communication barriers by harnessing the transformative power of technology.

This system employs a combination of speech recognition and image processing and NLP techniques to transform spoken language into text and subsequently into sign language output. The proposed solution [2] provides a practical and effective means of communication and interaction for the community facing hearing challenges.

The system [2] is specifically crafted to identify voice input using PyAudio, Sphinx, and the Google speech recognition API. After converting voice input into text, the output is then represented in sign language on the machine's screen, presented as a sequence of images or a moving video.

The system [2] employs a NLP pipeline that begins with the accurate transcription of voice input using technologies like PyAudio, Sphinx, and the Google speech recognition API. Once the spoken language is transcribed into text, NLP techniques come analyze the linguistic structure, semantics, and contextual nuances of the input.This process enables the system to comprehend the intended meaning and context behind the spoken words. The NLP model then translates the transcribed text into a representation suitable for sign language, considering grammatical rules and syntactic structures specific to sign languages. The ultimate result manifests a visual representation of sign language, showcased through either a sequence of images or an animated video on the screen. This presentation offers an efficient communication method for individuals facing challenges with hearing and speech, addressing their unique needs.

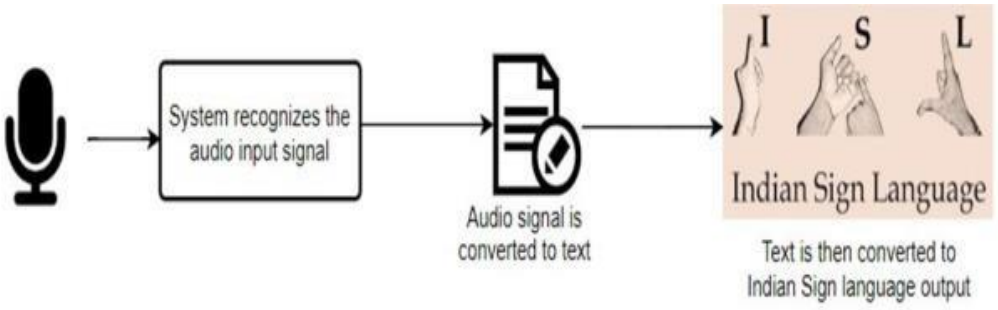


Figure 2: The architecture Diagram of Audio to Sign Language Converter

C. ACOUSTIC MODEL

Sign language plays a crucial role in fostering communication between the deaf and mute communities and the broader society, serving as a cornerstone of their cultural identity. It provides an inclusive platform, allowing Deaf individuals to actively engage, express themselves, and participate in society. However, effective sign language communication requires mutual language knowledge, which is not always guaranteed. This prototype aims to assess the feasibility of recognizing sign language, providing a means for hearing individuals to communicate with the deaf or mute by converting text into sign language images.

[3] The system begins by collecting a corpus, consisting of 1000 sentences on general information, used for training and testing. It undergoes training with 1000 sentences and testing with 150 sentences, and the knowledge is transferred to the language model. The language model comprehends the frequency of word co-occurrence in sentences, aiding the system in predicting the next words in a sentence. Utilizing a dictionary that associates words with their pronunciations, devoid of special symbols, the system employs an acoustic model.[3]This model functions as a detailed map, delineating how each sound in a word is spoken, derived from the analysis of recordings to identify sound patterns. To teach the system a new language, the sphinx training tool is utilized. This tool imparts knowledge about the patterns and sounds of words in the specified language, creating the necessary files for speech recognition. In essence, the system learns from example sentences, deciphers word sounds, understands how words fit together in sentences, and utilizes this acquired knowledge to comprehend spoken words

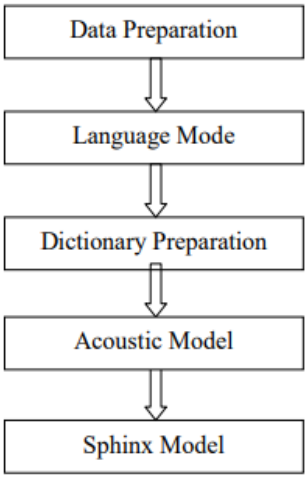


Figure 3: Block diagram of conversion system

D. HAMBURG NOTATION SYSTEM

In a world where spoken language predominates, individuals with hearing impairments encounter a significant disparity in their ability to express thoughts, emotions, and ideas. The inherent privilege of articulating oneself through spoken words is often taken for granted, leaving those reliant on sign language at a distinct disadvantage in various facets of everyday life. Recognizing this inequality underscores the critical necessity for the development of a sign language conversion system. Such a system would function as a vital link, enriching communication for those who depend on sign language and addressing the obstacles they face in a society primarily oriented toward spoken language.

The system provides an interactive solution for the hearing-impaired, converting English speech into 3D avatar animations depicting Indian Sign Language (ISL) signs. [4] Through the integration of Natural Language Processing (NLP) and a sign language database, the system becomes a valuable tool for communication among the hearing-impaired community. It encompasses a detailed presentation of the system's architecture, algorithms, and evaluation metrics, including experimental results highlighting accuracy, processing time, and memory utilization. Future efforts aim to improve the model by implementing a custom speech recognition system, replacing the existing Google API.

The system pioneers a robust approach in converting English speech into Indian Sign Language (ISL) animations. Drawing on the capabilities of Natural Language Processing (NLP), the Google Cloud Speech Recognizer API, and a predefined sign language database, this model operates seamlessly through its distinct components. Through speech-to-text conversion and input parsing in its initial phase, the system interprets [4] spoken language. Subsequently, the data preprocessing, ISL generation, and integration of HamNoSys notation in the second phase contribute to the accurate representation of sign language movements. The final stage involves SIGML notation and graphic generation, culminating in the creation of animated content. The utilization of HamNoSys notation proves instrumental in capturing the nuances of sign language, converted into SIGML files for animation synthesis. Beyond its technical capabilities, the ES2ISL system significantly reduces processing time, positioning it as a valuable tool for advancing communication and information exchange for individuals with hearing impairments.

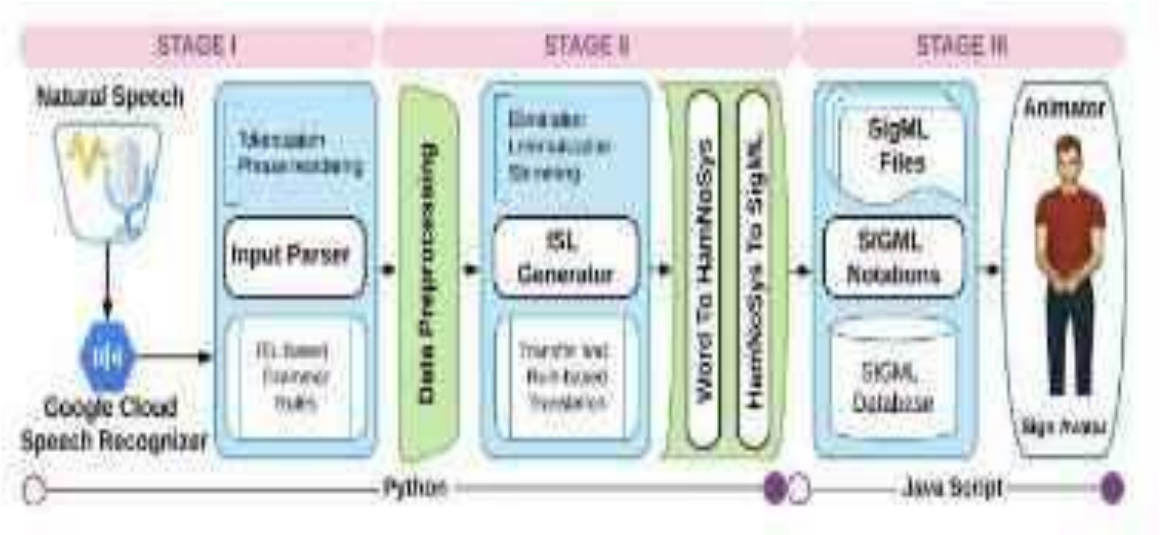


Figure 4: Architecture of speech to sign language conversion

III. PROPOSED SYSTEM

A. Architecture

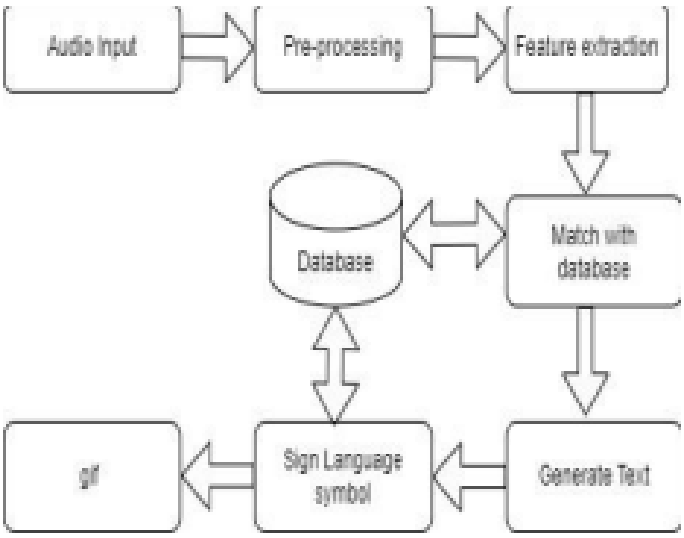


Figure 5: Proposed architecture

B. Implementation

In this proposed model of audio to sign language translation, the input for the system is taken from the user.The input audio then undergoes the preprocessing step. The preprocessing step involves cleaning and conversion of input into NumPy array. After the preprocessing, the important features of the input audio isextracted. By extracting relevant features, it becomes easier to train machine learning models, classify audio, and perform various tasks related to audio to sign language translation. The input audio the undergoes hyperparameter tuning. It optimizes the size of extracted features.

The input audio is then classified as train-test model to improve the accuracy and performance of the system. Different Machine Learning models are used to train and test the input audio and the model that obtains the highest accuracy is used to generate the output. After training, the input is then matched with the Database and the corresponding text is generated. If the text matches with audio signal in database, a corresponding gif of the sign language is produced.

IV. COMPARATIVE STUDY

The survey investigated various techniques for converting audio into sign language and found that employing a blend of diverse machine learning models yields significant effectiveness in this conversion process. The proposed system seeks to enhance the existing model by transitioning from generating sign language for individual letters to generating sign language for entire words with improved accuracy. Table I presents a comparative analysis of different models, shedding light on the methodologies that could be harnessed for this purpose. Leveraging a diverse technological toolkit, the survey integrates sophisticated methodologies, notably Neural Machine Translation (NMT) [5] and computervision[6]. These advanced technologies contribute to a thorough investigation of audio to sign language conversion processes, enhancing the overall comprehension

and depth of the subject matter. Furthermore, the integration of these methodologies ensures a robust and versatile framework for analysing the intricacies of sign language interpretation in audio formats.

METHODS	Advantages	Disadvantages
Convolutional Neural Network	Robust Features End –to –End Learning Emotion Recognition	Complexity Data dependency
Natural Language Processing Methods	Innovative solution Easy communication User-friendly interface	Limited dataset Lack of consideration of ethical issues
Acoustic Model	cost-effective easy to use	Does not currently incorporate facial expressions. The system is limited to Indian Sign Language
HamNoSys	Scalability Innovative Technology Multilingual Support	Limited Dataset Privacy Concerns
Neural Machine Translation	Provides a real time solution for translating Arabic Semantic Understanding	Limited only to Arabic Sign Language Small dataset is used for training
Computer Vision	Uses a Multi-tasking Strategy. Capture facial muscle movements	Errors in skeletal pose extraction. Poses challenges in precise interpretation.

Table 1:Comparative Study

V. CONCLUSION

This study underscores the myriad advantages inherent in real-time audio to sign language translation systems. Beyond significantly enhancing communication inclusivity for those with hearing impairments, these technologies facilitate seamless interaction across diverse contexts. Their remarkable speed and precision, driven by ongoing advancements in machine learning algorithms, empower users in effortless engagement. Moreover, these systems bolster accessibility, fostering social integration, expanding educational horizons, and supporting professional growth within deaf and hard-of-hearing communities. The overall positive impact on quality of life, combined with the potential for widespread adoption, underscores the transformative capacity of real-time audio to sign language translation technology in fostering a more inclusive and interconnected global society

VI. REFERENCES

[1] Sreeraksha M R, Vani H Y, Phani Bhushan, D K Shivkumar (August 2021), “Speech to Sign Language Conversion using Convolutional Neural Networks”, International Journal of Engineering Research & Technology (IJERT), Vol.09 Issue 12

[2] Rishin Tiwari, Saloni Birthare, Mr. Mayank Lovanshi, "Audio to Sign Language Converter"(2022), International Journal for Research in Applied Science & Engineering Technology (IJRASET), https://doi.org/10.22214/ijraset.2022.47271

[3] B. Sivaranjani, J. Sathya Priya, E. Ghanishka, V. Kamal Raj, Mohana Priya (June 2020), "Audio to Sign Language Translator Using Python" International Journal of Scientific Engineering and Applied Science (IJSEAS), https://ijseas.com/volume6/v6i6/ijseas20200601

[4] Bhavinkumar Devendrabhai Patel, Harshit Balvantrai Patel, Manthan Ashok Khanvilkar, Nidhi Rajendrakumar Patel, Thangarajah Akilan(2020), "ES2ISL: An Advancement in Speech to Sign Language Translation using 3D Avatar Animator", IEEE Canadian Conference on Electrical and Computer Engineering(CCECE),https://doi.org/10.1109/CCECE47787.2020.9255783

[5] Mounika Kanakanti, Shantanu Singh,Manish Shrivastava(2023),"MultiFacet: A Multi-Tasking Framework for Speech-to-Sign Language Generation",https://dl.acm.org/doi/pdf/10.1145/3610661.3616550

[6] Diana T. Mosa,Nada A. Nasef1,Mohamed A. Lotfy,Amr A. Abohany,Reham M. Essa, Ahmed Salem, "A real time Arabic avatar for deaf–mute community using attention", Neural Computing and Applications(2023)35:21709 21723,https://doi.org/10.1007/s00521-023-08858-6

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

LungCare AI : Pioneering Advancements in Respiratory Care using Deep Learning

Dr. Sarika S[1], Adarsh Jin[2], Allan T Jose[2], Arun Jibi[2], Emmanuel Jose[2]
*[1]Associate Professor, Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Vazhakulam, Ernakulam, Kerala, India
[2]Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Vazhakulam, Ernakulam, Kerala, India

ABSTRACT
The LungCare AI system, tailored for X-ray and CT scan image processing, employs diverse enhancement techniques to mitigate noise and enhance contrast. Following the pre- processing phase, a pre-trained Convolutional Neural Network (CNN) model is employed to analyze features like nodule shape, size, and distribution pattern. Cross-referencing with a comprehensive knowledge base ensures accurate identification, subject to verification by medical professionals, enabling prompt interventions to potentially avert disease progression. The proposed architecture streamlines appointment booking and image uploading, ensuring a secure workflow overseen by authorized personnel with a focus on prioritizing medical data confidentiality. Leveraging advanced technology for precise disease assessment contributes to enhanced patient care.
In tandem, traditional lung disease detection methods include skin tests, blood tests, and imaging techniques like chest X- rays and CT scans [8]. Recent advances in deep learning have transformed medical image analysis, particularly in diagnosing lung diseases. Inspired by the brain's structure, this sub-field of machine learning [13] excels in identifying patterns without hand-designed features, significantly improving performance in medical applications for efficient detection and classification of conditions. The transformative impact of the "LungCare AI" application, harmonizing innovative technology and user-centric design, signifies a positive shift towards improved respiratory health outcomes. This convergence underscores the transformative role of artificial intelligence in healthcare diagnostics.
Keywords: CNN, Deep Learning, Chest X-Ray, CT scan

I. INTRODUCTION

In the contemporary healthcare environment, the urgency to address respiratory challenges has reached unprecedented levels, necessitating a profound examination of transformative technologies. This survey meticulously explores the escalating need for advancements, with a specific focus on the early detection and management of diseases affecting lungs. Leading this technological revolution are advanced Deep Learning (DL) models, distinguished for their proficiency in deciphering subtle patterns indicative of respiratory conditions.

The survey endeavours to illuminate the pivotal role played by these advancements in reshaping the entire landscape of respiratory healthcare.

Beyond the realms of diagnostics, the survey recognizes that comprehensive care demands a holistic approach. To this end, it underscores the imperative integration of features such as appointment booking and direct chat communication with healthcare professionals. In an era where patient engagement and proactive management stand as paramount objectives, the incorporation of these user-centric elements becomes crucial. By shedding light on the symbiotic relationship between advanced DL technology and patient-oriented functionalities, the survey envisions not only improved healthcare outcomes but also a heightened quality of life for individuals navigating the intricate challenges posed by respiratory health issues. This holistic perspective lays the groundwork for a transformative paradigm, fostering collaborative and informed relationships between patients and healthcare providers, ultimately contributing to an elevated standard of care and well-being.

II. LITERATURE SURVEY

A. ECHO STATE NETWORK MODEL

In the past few years, notable advancements have been achieved in the realm of Artificial Intelligence (AI), particularly in the domain of biomedical diagnostics, such as the detection of cancer. This study

[1] introduces an innovative model by Harnessing the power of Deep Learning. The study employs Gabor filtering for image preprocessing and utilizes GhostNet for feature extraction. Hyperparameter adjustments through AFAO and cancer detection via TSA [6] with echo state network (ESN) contribute to the model's effectiveness. Extensive experimental results showcase the superior performance of BICLCD-TSADL, achieving a remarkable accuracy of 99.33%, signifying a noteworthy advancement in cancer detection efficiency.

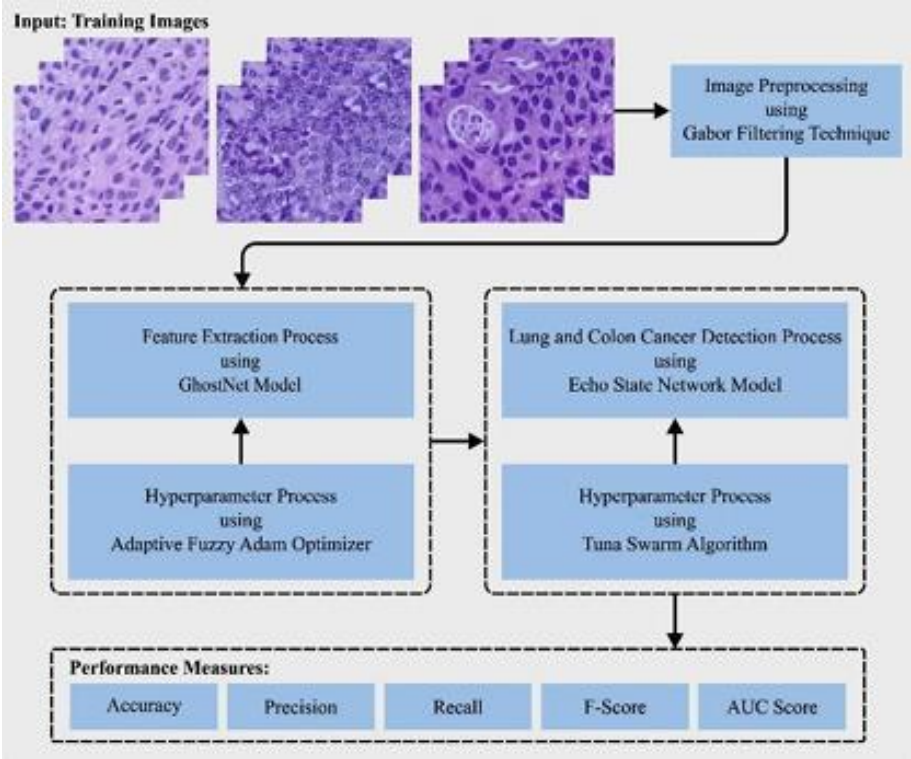


Fig. 1: Architecture Diagram

1) *Feature Extraction:* The GhostNet model is incorporated for efficient feature extraction. This innovative approach reduces computations and inputs, enhancing the overall effectiveness of the GhostNet model [1]. Central to GhostNet is the Ghost bottle- neck, a pivotal component that employs two ghost models to create mapping features. The process involves downsizing input mapping features through a typical convolution, followed by a linear operation on the downscaled features, generating a substantial quantity of ghost features. These outcomes are then amalgamated to produce an output mapping feature. To optimize the GhostNet model's performance, the Adaptive Firefly Algorithm Optimization (AFAO) is employed. The Adaptive Firefly Algorithm Optimization, modified iteration of stochastic gradient descent (SGD), enhances efficacy and expedites the learning process by iteratively evaluating biased and bias- corrected first and second moments, adjusting the model's parameters. The prediction accuracy of GhostNet with AFAO is subsequently evaluated using the testing set. This optimized GhostNet feature extractor, coupled with AFAO, improves the precise representation of colon and lung cancer image data, contributing to the improved overall performance of the biomedical image analysis system.
2) *Classification:* In the process of cancer detection and classification, the Echo State Network (ESN) model is utilized due to its effectiveness in training Recurrent Neural Networks (RNNs). This ensures streamlined training procedures and reduced time consumption compared to conventional statistical and machine learning methods. The ESN model computes reservoir layer updates using specific equations, and its output is determined through a weighted connection matrix. The optimal solution for the weighted matrix is obtained through the least squares or Mean Squared Error (MSE) method, contributing to an effective and time-efficient cancer detection system.

To achieve optimal parameter tuning in the ESN model, the study employs the Tuna Swarm Algorithm (TSA). TSA utilizes a spiral foraging strategy inspired by tuna schools, adapting its approach based on probability allocation and incorporating a fitness function derived from cancer detection outcomes. This approach enhances global exploration abilities and contributes to the overall effectiveness of the cancer detection system. Fig 1 shows the architecture diagram.

3) *Experiment Analysis:* The experimental validation of the proposed model in the study was conducted using Python 3.6.5 on a PC equipped with an i5- 8600k processor, GeForce 1050Ti 4GB GPU, 16GB RAM, 250GB SSD, and 1TB HDD. The model was simulated with specific parameter settings, including a learning rate of 0.01, dropout of 0.5, batch size of 5, 50 epochs, and ReLU activation.

The validation was performed on the dataset LC25000, comprising five classes, each consisting of 5000 samples. The approach demonstrated accurate identification and classification of all five class labels, as illustrated in confusion matrices, precision-recall analysis, and ROC investigation.

Detailed results on 70% of training (TRP) and 30% of testing (TSP) data demonstrated elevated accuracy, precision, F-score, recall, and AUC scores for each class.

Comparisons with other models revealed the superior performance of BICLCD-TSADL, emphasizing its effectiveness in medical image analysis. The BICLCD-TSADL model exhibited performance of 99.17% accuracy, 97.38% precision, 98.64% recall, 98.01% F-score, and an AUC score of 98.98%. The incorporation of IAFO and TSA algorithms for hyperparameter optimization further enhanced the model's overall outcomes, confirming its improved performance compared to existing techniques.

B. MULTI VIEW - KNOWLEDGE BASED COLLABORATIVE MODEL

The research [2], classifies lung nodules as benign or malignant on chest CT scans [7]. The model decomposes the three-dimensional image of each lung nodule into 9 fixed plane views and uses a knowledge-based collaborative approach. It employs three ResNet-50 neural networks to understand three basic properties of a nodule - Appearance, Shape and Voxel heterogeneity and with help from a penalty loss function controls the balance between false negative and false positive rates.

The methodology of the said algorithm comprises of four major steps: extraction of two dimensional nodule slices from the nine view planes, extracting patches representing overall appearance of the lung condition, heterogeneity in voxel values, and shapes on two dimensional nodule slices, constructing nine submodels and training each of them using the patches extracted on each view of planes, and con- struction and training of the model for classification. The model's performance is evaluated using the database termed LIDC-IDRI, and it achieved superior accuracy in comparison to state-of-the-art approaches. The study also compared the model's performance with other deep learning methods and traditional CADs (Computer-Aided Diagnostic systems), demonstrating the effectiveness of the multi- view learning approach.

In summary, the MV-KBC model uses a multi-view approach and a penalty loss function to accurately seggregate lung nodules as non-cancerous or cancerous, showcasing the potential of deep learning for medical imaging and contributing to the detection of lung cancer before its obvious onset.

1) *Multi-View Slice Extraction:* The multi-view slice extraction method used in the research involves decomposing three-dimensional lung nodules into nine fixed view planes, namely coronal, axial, sagittal, and six diagonal planes. From these views, two-dimensional nodule slices are extracted, and overall appearance (OA), heterogeneity in voxel values (HVV), and shapes (HS) patches are further extracted. These patches are used to train a submodel for each view, where three pre-trained ResNet-50 neural networks are fine-tuned to describe the OA, HVV, and HS of the lung nodules. The submodels are then combined to construct and train the complete model for lung nodule classification. This multi-view approach aims to leverage complementary information from different perspectives to improve the accuracy of lung nodule classification. The method is illustrated in Fig. 2. The approach aims to capture the 3D characteristics of the nodules from multiple perspectives, enhancing the model's ability to accurately classify lung nodules as non-cancerous or cancerous.

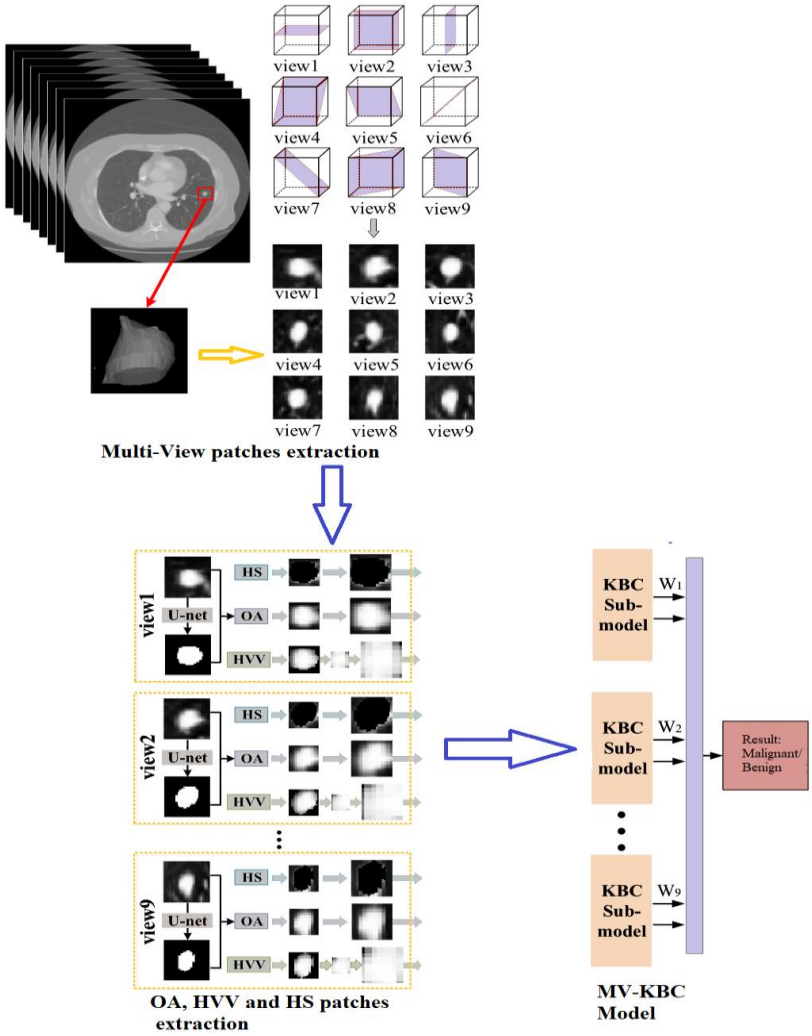


Fig 2: Framework of the proposed algorithm

2) *OA, HVV and HS Patches Extraction:* The method of Extracting OA, HVV, and HS Patch involves decomposing 3D lung nodules into nine fixed view planes and extracting 2D nodule slices from each view. The U-Net network is employed for nodule segmentation, and the extracted patches are used to train a submodel. Each submodel contains three pre- trained ResNet-50 neural networks, which are finetuned to characterize the overall appearance (OA), heterogeneity in voxel values (HVV), and heterogeneity in shapes (HS) of the lung nodules. The extracted patches are used to train the submodel, and the complete model is constructed and trained for lung nodule classification. Data augmentation is applied to alleviate overfitting, and the model's performance is evaluated using various metrics like sensitivity, accuracy, etc. The method aims to lever- age the complementary information from different perspectives to improve the accuracy of lung nodule classification.

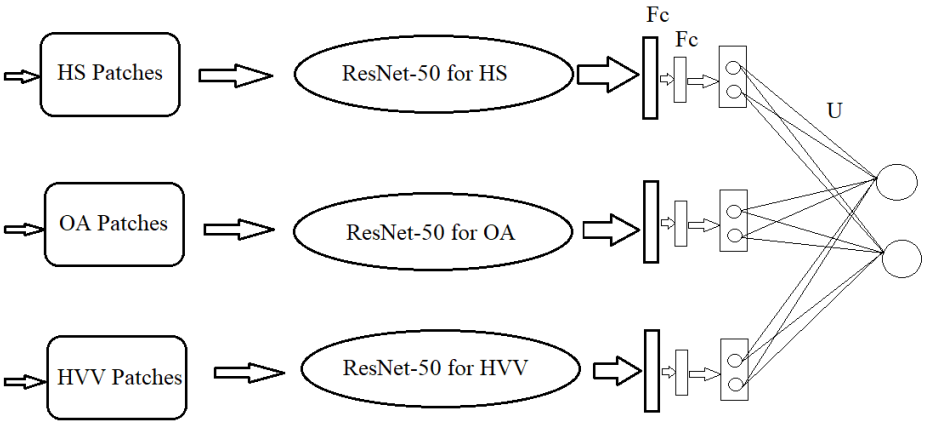


Fig 3: Architecture of a submodel for a specific view

3) *KBC Submodel:* The proposed submodel is designed to describe lung nodules from different perspectives, namely overall appearance (OA), heterogeneity of voxel values (HVV), and heterogeneity in shapes (HS). The submodel consists of three pre- trained ResNet-50 networks, each fine-tuned to capture specific nodule characteristics. The architecture of a submodel involves extracting OA patches, HVV patches, and HS patches from two- dimensional nodule slices on each of the nine fixed views of planes. These patches are used to train each of the submodels, and the outputs of the three ResNet-50 networks are combined to form the prediction made by the complete model. The KBC sub-model aims to leverage the complementary features captured by the ResNet-50 networks from different perspectives, enhancing the model's ability to accurately classify lung nodules as benign or malignant. The submodel's architecture is illustrated in Fig 3 of the paper.

Fig.3 illustrates the architecture of the Knowledge- Based Collaborative (KBC) sub-model for lung nodule classification. It shows three sets of patches representing different nodule characteristics: heterogeneity in shapes (HS), overall appearance (OA), and heterogeneity in voxel values (HVV). Each set of patches is processed by a dedicated ResNet-50 network, designed to extract relevant features for that characteristic. After feature extraction, the outputs from the three networks are passed through fully connected (Fc) layers and then integrated in a collaborative unit (U), which combines the features to make a final classification decision. The figure visually represents the parallel processing of different nodule features and the subsequent integration for diagnosis.

4) *MV-KBC Model:* The model, or the complete model comprises of nine KBC sub-models, employs a unique architecture. Each submodel features a double neuron output layer linked to a common single neuron in the classification layer, followed by the sigmoid function. The overall model prediction is derived from the output of this classification layer.

To address the potential cost imbalance of misclassifications in medical contexts, where misidentifying a malignant nodule as 'benign' or non-cancerous may have graver consequences than the reverse, the authors propose a penalty cross-entropy loss. Unlike traditional cross-entropy loss, this variant allows for differential penalization of false negative and false positive errors. In medical scenarios, such as lung tumor detection, this approach aims to mitigate the risk of overlooking early-stage tumors by discouraging the false reassurance of benign classifications.

5) *Performance Analysis:* The MV-KBC algorithm proposed here has achieved 91.60% Accuracy, 86.52% Sensitivity, and 87.75% Precision in classifying lung nodules. It also received an F1 score of 87.13% and thus has proved to be a viable tool in the domain.

The emphasis lies on the enhanced performance in distinguishing nodules within each Median Malignancy Level (MML) subgroup, achieved through the utilization of a multi-view architecture and penalty loss that regulates the balance between false negative and false positive rates. However, it is crucial to acknowledge the potential risk of overfitting in the deep model employed, stemming from insufficient training data.

C. U-NET AND CUSTOM IRV2 INTEGRATION MODEL

A novel Computer-Aided Diagnostic (CAD) system is proposed in [3] to empower radiologists with enhanced accuracy and efficiency in diagnosing pulmonary diseases from chest X-rays [7] (CXRs). The system leverages a fine-tuned Convolutional Neural Network (CNN) architecture, analyzing CXR images [5] in two stages: initial healthy/infected classification followed by in-depth disease type identification for confirmed infections. Lung region segmentation ensures focus on relevant information, boosting processing efficiency and noise reduction. This approach surpasses existing methods in both segmentation and classification, as validated on the standard NIH chest X-ray dataset (detailed architecture visualization omitted).

1) *Segmentation:* This study [3] highlights the crucial role of lung segmentation in enhancing the ac- curacy of pulmonary disease classification in chest X-rays. Segmenting the lung region offers several advantages: it reduces computational demands by focusing on relevant areas, concentrates analysis on features directly associated with lung diseases, and minimizes interference from irrelevant background information. To achieve this segmentation, the re- searchers utilized a U-Net architecture, characterized by its distinctive encoder-decoder structure. The contraction path of the proposed model entails a series of iterative 3x3 convolution operations, succeeded by Rectified Linear Unit (ReLU) Activation, and 2x2 max pooling to achieve down sampling. At each operation in this path, there is a twofold increase in feature channels, contributing to the overall efficiency and representational capacity of the model. The expansion phase of the model relies on up-convolution layers, combined with skip connections that retrieve lost detail from earlier stages. Training is driven by manually segmented lung masks extracted from the NIH Chest X-ray dataset.

Training spans a maximum of 200 epochs, with hyperparameters fine-tuned for optimal performance. Dropout (50%) is introduced for regularization in the expanding path, and batch normalization enhances efficiency and stability. Hyperparameters are adjusted considering computational constraints, with specific values such as a batch size of 32, a U- Net depth of 5, and two convolutional operations per depth. The model attains optimal performance following these adjustments.

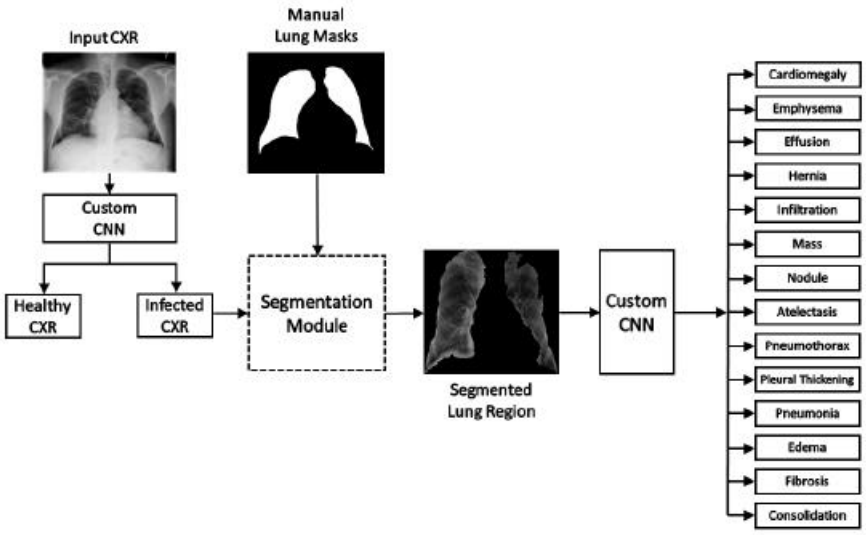


Fig 4: Architecture Diagram

2) *Classification:* This study tackles the issue of low accuracy in classifying specific pulmonary dis- eases by proposing a custom CNN architecture that extends Inception-ResNet. The model first distinguishes CXR images as healthy or infected, followed by a segmentation module for isolating the lung region in infected images. The segmented region undergoes a secondary evaluation to ex- amine disease-specific texture and shape features at a micro-level for a multi-class classification. In this process, the custom Convolutional Neural Network (CNN) incorporates four supplementary convolutional layers before reaching the Inception- ResNet, ensuring the preservation of image resolution to facilitate micro-level feature extraction. Hyperparameters, including learning rate, optimizer, and activation function, are fine-tuned for optimal performance.

The optimizer algorithm used is Adam. Inception- ResNet extracts meaningful features, and the output traverses additional fully connected layers for weight determination through backpropagation. The classification layer initially handles binary classification and is later adapted for multi-class classification of 14 pulmonary diseases, resulting in enhanced overall accuracy and individual class accuracies. Fig 4 shows the architecture diagram.

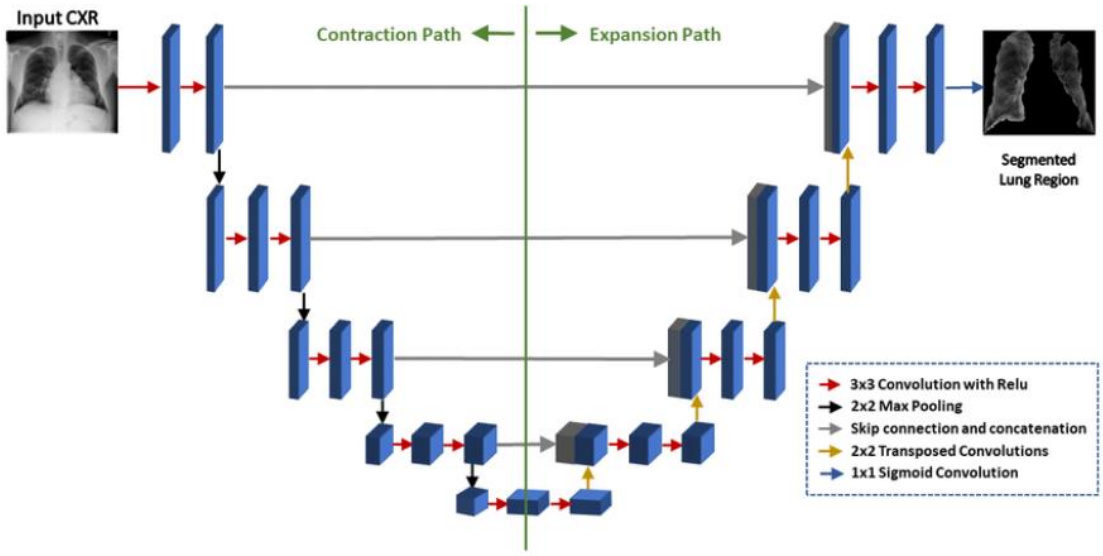


Fig 5: U-Net Architecture

3) *Experimentation and Results:* To streamline the computationally demanding task of training a CNN, a GTX 1080ti GPU with 8 GB memory is combined with a standard system featuring a core i7 8th Gen CPU and 16 GB system memory. Code implementation is carried out in Python using Keras with TensorFlow as the backend engine. The training process involved 200 epochs for both segmentation and classification tasks. Data was carefully split: 70% for training, 10% for validation, and 20% for testing, ensuring no patient data overlapped between sets. The dataset utilized comprised 112,120 images, with 78,484 specifically used for classification training. A batch size of 64 images resulted in 1,227 steps per epoch across the 200 training epochs.

Building upon a modified U-Net architecture, researchers pushed the boundaries of lung segmentation and disease classification in chest X-rays [11]. Leveraging pre-trained weights and the vast NIH dataset, their model achieved state-of-the-art performance, boasting a remarkable mean Dice score of 0.9734 and IoU of 0.9646 for lung segmentation. While specific details of the architecture and training regimen are restricted, the results indicate significant promise for this approach in advancing medical image analysis. Strikingly, it distinguishes healthy and infected CXR images with 91% accuracy and exhibits exceptional class-wise accuracies between 0.975 and 0.984 for 14 distinct pulmonary diseases. Notably, the entire CXR image is used for healthy vs. infected classification, while for infected cases, the model focuses on the segmented lung region for deeper analysis. These remarkable findings suggest the proposed U-Net architecture's potential to significantly improve medical professionals' capabilities in lung analysis and diagnosis. This process, involving the collection and utilization of manual lung masks, stands as a notable contribution of this research study.

D. CNN MODEL-BASED LUNG DISEASE IDENTIFICATION

Lung diseases such as Lung Cancer, Pneumonia and COVID-19[10] are the most found diseases in human beings. Lung diseases [9] need to be diagnosed timely. To serve this purpose many machine learning models have been developed. Vanilla network, capsule network, and VGG net algorithms serve this purpose. In the study [4] Convolutional Neural Network is used for predicting lung diseases from Chest X-Ray [11] images sourced from the Kaggle repository dataset. Spyder, Keras and TensorFlow are the tools used for implementation. 93% of mean accuracy is yielded by model. Diseases which include lung cancer, Pneumonia, covid or none are predicted by this model.

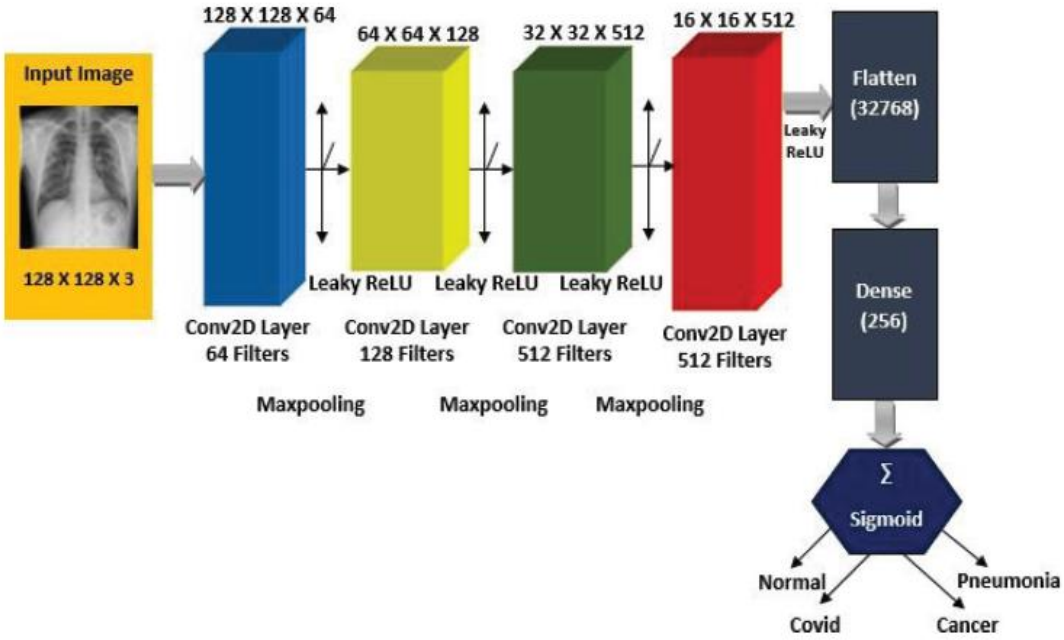


Fig 6: Architecture Diagram

1) *Proposed Model:* Convolutional Neural Network: The primary task of CNN [12] is classification, offering a scalable approach not only for classification but also for recognition of objects. CNN is further divided into three layers.
2) *Convolutional Layer:* The Convolution layer is the cornerstone of CNN, where most of the computation takes place. It plays a crucial role in determining complex data representations that can be visualized through extensive datasets. The fundamental inspiration for convolutional neural networks is drawn from the organization of visual cortex in animals, the regions responsive to different parts of the visible field.

In this context, neuronal cells exhibit responses to the presence of edges in defined orientations. For instance, some cells activate when exposed to horizontal edges, while others respond to vertical edges or diagonal edges. The convolution operation involves using an input image with a filter of M×M size in this layer. This filter moves across the input image, and dot product is performed with the specified size of M×M. The resulting output called feature maps provides information on edges and corners.

3) *Pooling Layer:* Pooling layer is after the Convolution Layer. The size of the feature map that gets convolved is reduced, this decreases the computational costs. To operate independently on each of the feature map, the connections in layers are reduced, serving as a bridge between the Convolution layer and the Fully Connected Layer. Various operations of pooling exist according to the method employed. Max Pooling selects the largest element from each section of the feature map. Average Pooling calculates the average of the elements within a predefined-sized image section and Sum Pooling determines the sum of the elements in that section.
4) *Fully Connected Layer:* This layer contains the weights and biases essential for connecting neurons between the two layers. The input from the previous layer is flattened and transmitted to the Fully Connected layer. Within this layer, the flattened vector undergoes various mathematical operations. The classification process takes place in this process. Fig 6 Shows the Architecture diagram.

5) *Activation Function:* Network Model. This method identifies all types of relationships between the variables in the network. This method introduces nonlinearity to the network. Various activation functions such as Leaky ReLU and Sigmoid are employed for this purpose, each serving different roles within the model. The Sigmoid function is employed in the Binary Classification CNN model and SoftMax is employed for the Multiclass Classification.

A Comparative study of the advantages and disadvantages of the methodologies proposed in the various reference papers are provided in Table 1.

III. PROPOSED SYSTEM

The devised LungCare AI system is engineered to process CT scan and X-Ray images of the lungs. Employing diverse enhancement techniques during preprocessing mitigates noise and enhances contrast. Subsequently, these refined images undergo analysis by a pre-trained Convolutional Neural Network (CNN) [12] model. Specific features are scrutinized by the CNN model, such as nodule shape, size, and distribution pattern, cross-referencing its findings with a comprehensive Knowledge base. Upon identifying relevant features, the model generates an output, subject to verification by a medical professional before communicating results to the patient. This expeditious diagnostic approach facilitates prompt initiation of remedial actions, potentially averting disease progression. The system's efficiency lies in its ability to swiftly and accurately deliver probable results, contributing to timely intervention and improved patient outcomes. Fig. 7 shows the proposed system architecture.

Patients can conveniently book appointments within the doctor's schedule, during which the doc- tor can upload medical images for preprocessing before model input. The system generates prediction results and a diagnostic report. Exclusively, the doctor forwards images to the admin for result generation, incorporating prediction details. This meticulous process ensures a secure and efficient workflow, prioritizing medical data confidentiality. Advanced technology facilitates precise disease assessment, contributing to enhanced patient care. The stream- lined process, overseen by authorized personnel, ensures confidentiality, while advanced technology enhances diagnosis of diseases, promoting accurate and timely medical interventions for improved patient care.

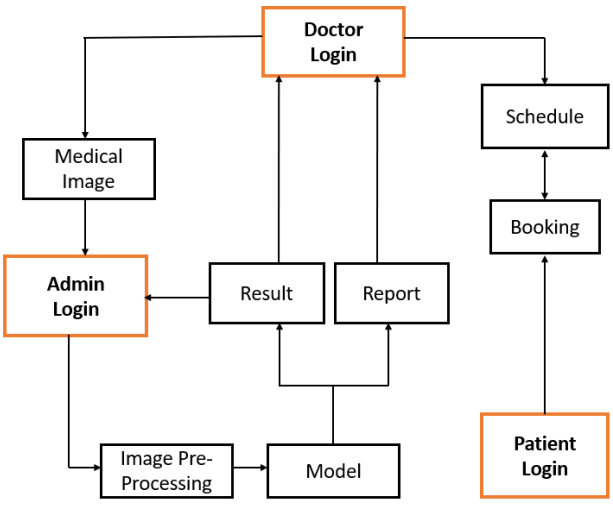


Fig 7: Proposed Architecture

IV. COMPARATIVE STUDY

Based on the reviewing of the papers, a thorough comparison analysis was conducted. This study attempted to identify and evaluate the different benefits and drawbacks that are present in every technology that was discussed in the literature. Table 1, which methodically provides a comprehensive summary of the observed contrasts and captures the complex subtleties of the technologies covered in the corresponding papers, is essential to this analysis. After a methodical examination and evaluation, the table provides significant value as a tool for comprehending the technologies under investigation and provides information about their individual advantages and disadvantages.

Title	Technique Used	Advantages	Disadvantages
Biomedical Image Analysis for Colon and Lung Cancer Detection Using Tuna Swarm Algorithm With Deep Learning Model	GhostNet and Ecostatic network	High Accuracy	Higher Complexity
Knowledge-based Collaborative Deep Learning for Benign-Malignant Lung Nodule Classification on Chest CT	MV-KBC approach	Multi-view learning, Collaborative Sub models	Nodule segmentation accuracy affect classification
Lung Segmentation-Based Pulmonary Disease Classification Using Deep Neural Networks	Image segmentation based approach	Focused feature extraction without loss of information	Manual annotation requirement and high computer resources
Lungs Diseases Prediction based on Convolutional Neural Network	CNN algorithm	Early Detection	Data limitations and overfitting

Table 1: Comparative study of the reference papers

V. CONCLUSION

In conclusion, we examine the" LungCareAI" application, highlighting its pivotal role in revolutionizing respiratory healthcare. Through the strategic fusion of innovative technology and user-centric design, the mobile health solution utilizes deep learning algorithms to swiftly and accurately predict lung diseases. The integration of features like appointment booking and direct communication with healthcare professionals enhances the user's experience. With robust report generation capabilities, the application becomes a valuable tool for users and healthcare providers. As a beacon of progress in mobile health," LungCareAI" signifies a positive shift towards improved respiratory health outcomes, showcasing the transformative impact of artificial intelligence in healthcare diagnostics.

VI. REFERENCES

[1] S. Urooj, S. Suchitra, L. Krishnasamy, N. Sharma and N. Pathak, "Stochastic Learning-Based Artificial Neural Network Model for an Automatic Tuberculosis Detection System Using Chest X-Ray Images," in IEEE Access, vol. 10, pp. 103632-103643, 2022, doi: 10.1109/ACCESS.2022.3208882.

[2] Y. Xie et al.," Knowledge-based Collaborative Deep Learning for Benign-Malignant Lung Nodule Classification on Chest CT," in IEEE Transactions on Medical Imaging, vol. 38, no. 4, pp. 991-1004, April 2019, doi: 10.1109/TMI.2018.2876510.

[3] S. Z. Y. Zaidi, M. U. Akram, A. Jameel and N. S. Alghamdi, "Lung Segmentation-Based Pulmonary Disease Classification Using Deep Neural Networks," in IEEE Access, vol. 9, pp. 125202-125214, 2021, doi: 10.1109/ACCESS.2021.3110904.

[4] S. Shirsat and S. V. Kedar, "Lungs Diseases Prediction based on Convolutional Neural Network," 2021 International Conference on Computing, Communication and Green Engineer- ing (CCGE), Pune, India, 2021, pp. 1-5, doi: 10.1109/CCGE50943.2021.9776371

[5] X. Zhang et al., "CXR-Net: A Multitask Deep Learning Network for explainable and Accurate Diagnosis of COVID-19 Pneumonia From Chest X-Ray Images," in IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 2, pp. 980-991, Feb. 2023, doi: 10.1109/JBHI.2022.3220813

[6] M. Obayya, M. A. Arasi, N. Alruwais, R. Alsini, A. Mohamed and I. Yaseen, "Biomedical Image Analysis for Colon and Lung Cancer Detection Using Tuna Swarm Algorithm With Deep Learning Model," in IEEE Access, vol. 11, pp. 94705-94712, 2023, doi:10.1109/ACCESS.2023.3309711 mages," in IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 2, pp. 980-991, Feb. 2023, doi: 10.1109/JBHI.2022.3220813

[7] P. Yadav, N. Menon, V. Ravi and S. Vishvanathan," Lung-GANs: Unsupervised Representation Learning for Lung Disease Classification Using Chest CT and X-Ray Images," in IEEE Transactions on Engineering Management, vol. 70, no. 8, pp. 2774-2786, Aug. 2023, doi: 10.1109/TEM.2021.3103334

[8] M. A. Alzubaidi, M. Otoom and H. Jaradat, "Comprehensive and Comparative Global and Local Feature Extraction Framework for Lung Cancer Detection Using CT Scan Images," in IEEE Access, vol. 9, pp.158140-158154, 2021, doi: 10.1109/ACCESS.2021.3129597

[9] E. A. A. Alaoui, S. C. K. Tekouabou, S. Hartini, Z. Rustam, H. Silkan and S. Agoujil, "Improvement in automated diagnosis of soft tissues tumors using machine learning," in Big Data Mining and Analytics, vol. 4, no. 1, pp. 33-46, March 2021, doi: 10.26599/BDMA.2020.9020023

[10] Malik H, Anees T, Din M, Naeem A. CDC-Net: multi-classification convolutional neural network model for detection of COVID-19, pneumothorax, pneumonia, lung Cancer, and tuberculosis using chest X-rays. Multimed Tools Appl. 2023;82(9):13855-13880. doi: 10.1007/s11042-022-13843-7

[11] Abbas, Asmaa&Abdelsamea, Mohammed &Gaber, Mohamed. (2021) Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. Applied Intelligence. 51. 1-11.10.1007/s10489-020-01829-7

[12] Punn NS, Agarwal S. Automated diagnosis of COVID-19 with limited posteroanterior chest X-ray images using fine-tuned deep neural networks. ApplIntell (Dordr). 2021;51(5):2689-2702. doi: 10.1007/s10489-020-01900-3

[13] S. Mukherjee and S. U. Bohra, "Lung Cancer Disease Diagnosis Using Machine Learning Approach," 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India, 2020, pp.207-211, doi: 10.1109/ICISS49785.2020.9315909

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

Unveiling the Power of LLMs in Education

Manjusha Mathew[1], Abel Mathew Xavier[2], James Antony Paul[2], Shan Shaji[2], Vishnu S[2]
*[1]Assistant Professor, Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Vazhakulam, Ernakulam, Kerala, India
[2]Department of Computer Science and Engineering Viswajyothi College of Engineering and Technology, Vazhakulam, Ernakulam, Kerala, India

ABSTRACT
The early stages of large language models (LLMs) show great potential for aiding various fields, particularly education. However, both educators and students encounter challenges in obtaining precise and accurate answers to academic questions. To address this issue, this paper proposes the development of a fine-tuned LLM [1] specifically tailored to students' syllabi. This fine-tuning process would be based on the content of the relevant textbooks and specific data related to the courses outlined in the syllabus.The potential of LLMs in education is significant, offering assistance in tasks such as answering questions [2], providing explanations, and generating educational content. Nevertheless, the generic nature of existing LLMs can lead to inaccuracies or irrelevant information when applied to specific academic contexts. By fine-tuning an LLM to align with the content and requirements of students' courses, it can be better equipped to provide accurate and contextually relevant information [2].This approach holds promise for enhancing the educational experience by enabling students to access tailored and dependable information aligned with their curriculum. Educators, in turn, can benefit from a tool that supports the learning process and provides students with accurate resources. Through the utilization of course-specific data and textbooks, the fine-tuned LLM can offer a more targeted and reliable resource for addressing academic inquiries [3].The proposal to develop a fine-tuned LLM tailored to students' syllabi presents an opportunity to address the challenge of obtaining accurate and contextually appropriate information in the educational setting [4]. By leveraging this approach, educators and students can potentially benefit from an LLM that is better aligned with the specific requirements of the curriculum.

Keywords: Large Language models, Named Entity Recognition, NLP, Question answering.

I. INTRODUCTION

In today's digital education landscape, students often struggle to efficiently grasp and retain information from a multitude of academic resources. Lengthy documents and textbooks can be overwhelming, impeding effective self study. Additionally, students face challenges in clarifying doubts and obtaining concise explanations for complex topics. To address these issues, there is a demand for a user friendly AI-powered educational platform. This platform would offer automated document summarization, question answering features, and seamless access to external resources, streamlining the learning process for students. The difficulty that students face is

the unavailability of reliable and authentic LLM models that can give precise answers to the questions or prompts they ask. Many LLM models give information that may not be relevant to the student but it may be needed in understanding the concept. Sometimes it may generate content that will not be applicable for the students' program. The intersection of Natural Language Processing (NLP) and education has given rise to innovative applications, transforming the traditional methods of knowledge dissemination. Within the expansive domain of computer science education, the integration of NLP presents a unique opportunity to reimagine and enhance the learning experience for both students and educators. This survey paper seeks to delve into the latest developments in NLP as applied to computer science engineering syllabi, with a specific focus on a cutting-edge project designed to leverage a large language model. The cornerstone of our project lies in its distinct training dataset, exclusively of computer science syllabus. Through the use of language models, our system boasts a range of sophisticated capabilities, including automated summarization of educational content, dynamic question answering functionalities, and the generation of tailored quizzes for educators. Technologically, our project relies on advanced NLP models, such as BERT architecture, to comprehend and analyze the intricacies of computer science syllabi. The incorporation of these models enables our system to generate coherent and contextually relevant summaries, answer inquiries posed by students, and craft quizzes that align with the specific learning objectives outlined in the syllabi. Moreover, the integration of machine learning algorithms facilitates continuous improvement and adaptation of the model based on user interactions, enhancing its responsiveness and effectiveness over time. As we embark on this survey, our goal is to offer a detailed exploration of the existing landscape of NLP applications in education. We will spotlight key advancements, elucidating their impact on learning outcomes and addressing the unique challenges associated with computer science education. By taking a closer look at the technologies underpinning these advancements, we aim to provide educators, researchers, and technologists with insights into the potential benefits and challenges associated with the integration of NLP tools into the computer science learning environment. This comprehensive understanding will contribute to ongoing discussions on optimizing NLP applications for computer science education, ensuring a dynamic and responsive learning experience for students.

II. LITERATURE SURVEY

A. MRC FRAMEWORK Machine Reading Comprehension

(MRC) [1] is the technique used to train computers to understand documents or data provided for the target domain. MRC Framework proposed in the paper can generate pseudo-questions and the corresponding pseudo-answers. This framework uses reinforced learning which rewards the model for generating accurate pseudo-question. The generated pseudoquestion/answer pair with better F1-score is then used as data to self-train the model. The proposed framework can answer questions that are related to the target domain from the data augmented self-training provided by pseudoquestion/answer.

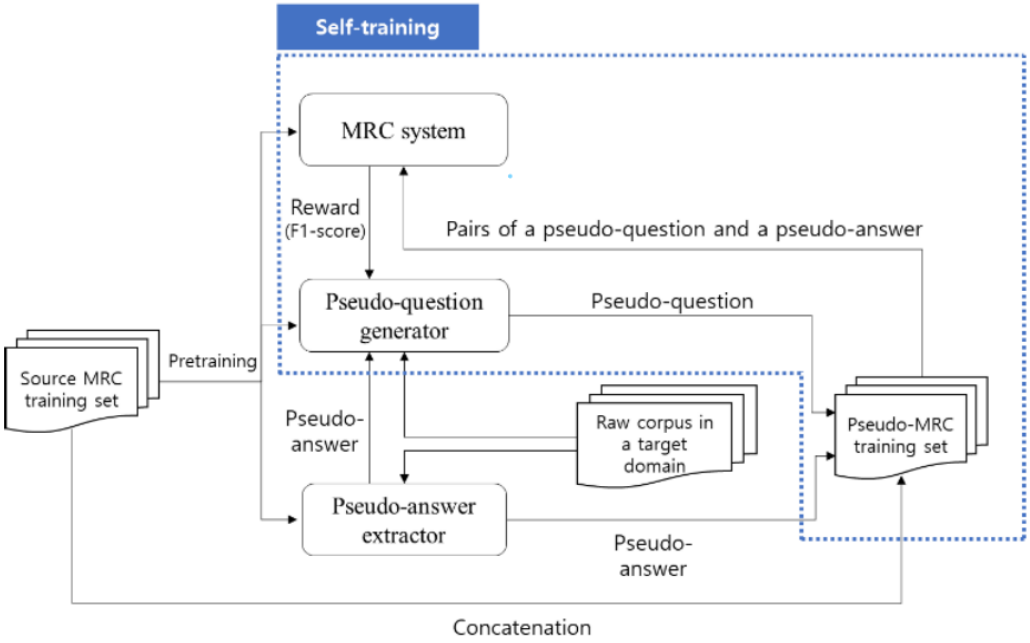


Fig 1. Architecture of proposed MRC framework

The framework, as depicted , consists of a pseudo-answer extractor, a pseudo-question generator, and an MRC system. The pseudo-answer extractor automatically identifies phrases that can be considered as correct answers, while the pseudo-question generator creates suitable questions for the extracted answers from the documents in the target domain [1]. It uses a modified pointer generator that can create questions that are contextually relevant to the pseudo-answers extracted from the documents.The proposed MRC system for domain adaptation includes a question generator and an answer extractor. The question generator is retrained using reinforcement learning and F1-scores as rewards to improve the quality of the generated questions. The MRC system is trained on a combination of labeled data from the source domain and pseudo-labeled data generated by the question generator and the answer extractor in the target domain. The system calculates rewards for reinforcement learning based on the F1-scores between the predicted answers and the answers provided. The self-training scheme based on this framework shows better performance than the system based on automatic data augmentation. The framework provides the domain adoption techniques that can improve MRC system performance in transition to a new domain. The implemented pseudo-question generator produces better results due to the modified pointer generator

B. MTA ENCODER FOR ANSWER GENERATION

The Multi-Layer Transformer Aggregation Encoder for Answer Generation (MTA) mentioned in [2] is a deep learning-based model that aims to improve the accuracy of answer generation tasks by enhancing the contextual correlation and semantic information features extraction. Contextual data at various layers is used by the MTA model, which is based on a multi-layer attention Transformer unit to model sequences. It can concentrate on information representation at different locations and combine nodes at the same layer to combine context information. It allows the integration of semantic information from the bottom layer to the upper layer, improving the encoder's information representation. The MTA encoder architecture comprises multiple layers of multi-head attention and feedforward neural networks. Each layer utilizes a self-attention

mechanism to capture semantic information at various sequence levels. Additionally, a novel position encoding method is introduced to enhance the relationship of each word by leveraging absolute and relative position information. This approach enables the attention mechanism to consider the positional order of the input, leading to improved performance. The MTA model has been assessed on the SQuAD dataset, achieving results with an EM score of 71.1 and an F1 score of 80.3. Ablation experiments have validated the effectiveness of each model component. Overall, the MTA model offers a promising solution to the constraints of traditional deep learning-based answer generation models by enhancing contextual correlation and semantic information extraction, with potential applications beyond answer generation tasks.

C. T-BERTSUM MODEL

The T-BERTSum model represents a significant advancement in the field of text summarization, particularly in its focus on topic-awareness and coherence. By leveraging Bidirectional Encoder Representations from Transformers (BERTs) and integrating a neural topic model (NTM), the model addresses the limitations of existing methods and offers a novel approach to generating coherent and informative summaries. This innovative framework demonstrates the potential to enhance the quality of text summarization by capturing essential information while minimizing redundancy. In the context of text summarization research, TBERTSum's emphasis on topic-awareness and coherence aligns with the broader challenges faced by existing methods. The model's two-stage extractive–abstractive framework represents a notable contribution to the field, aiming to ensure the relevance and coherence of generated summaries. Furthermore, the model's performance evaluation on the CNN/Daily Mail dataset showcases its results in terms of ROUGE scores and human evaluation metrics, underscoring its potential as a promising solution to the challenges of producing high-quality, topic-aware summaries with reduced redundancy. Overall, T-BERTSum's innovative approach to topic-aware text summarization presents a significant advancement in the field, offering a valuable contribution to the ongoing efforts to improve the coherence and informativeness of generated summaries. The model's successful performance on benchmark datasets underscores its potential to address the limitations of existing methods and highlights its relevance within the broader landscape of text summarization research.

D. BERT: SELF-SUPERVISED ATTENTION

The BERT model, short for Bidirectional Encoder Representations from Transformers, has shown impressive performance across various natural language processing tasks. However, when fine-tuned on smaller datasets, BERT models often suffer from overfitting, leading to degraded performance. To overcome this challenge, "Improving BERT With Self-Supervised Attention" introduces a new method called Self-Supervised Attention (SSA). The Self-Supervised Attention (SSA) [4] mechanism enhances the BERT model's attention layer by integrating auxiliary knowledge, without the need for extra data or annotations. SSA iteratively generates weak, token-level attention labels, enabling the model to learn more robust representations and reduce overfitting on smaller datasets. By doing so, SSA helps the model capture the relationship between tokens and create a smoother surface for optimization, ultimately improving the model's generalization capabilities. The effectiveness of the self supervised attention mechanism is demonstrated through empirical evaluations on sentiment analysis tasks and the General Language Understanding Evaluation (GLUE) benchmark, highlighting its ability to enhance BERT's generalization capabilities.

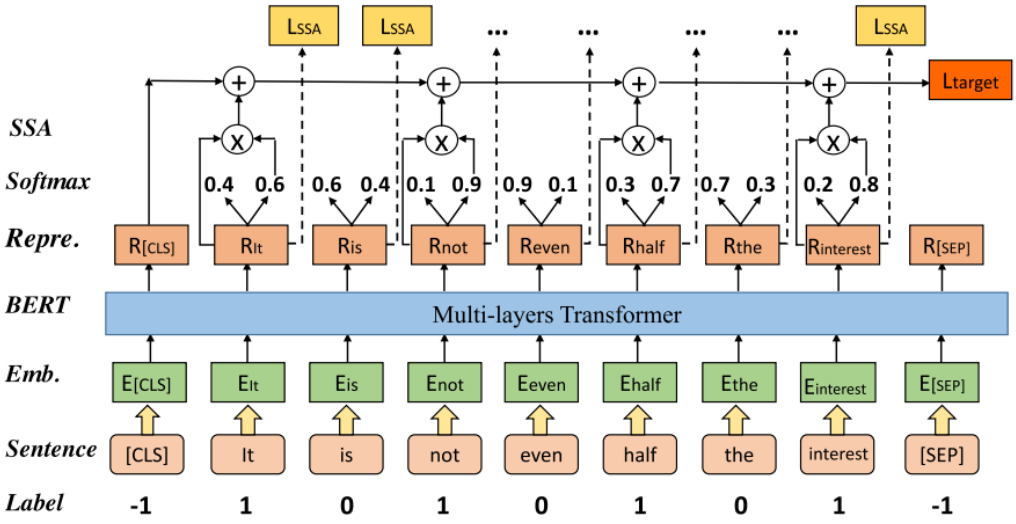


Fig 4. Illustration of an example based on the hybrid model.

Fig 4. Illustration of an example based on the hybrid model. The integration of Self-Supervised Attention (SSA) into BERT is explored in two different ways, and a hybrid approach is proposed to combine their advantages. Empirical evaluations on various public datasets demonstrate significant performance improvements using the SSA-enhanced BERT model. The proposed SSAHybrid model outperforms the base BERT model and other existing works on sentiment analysis tasks and the General Language Understanding Evaluation (GLUE) benchmark, showcasing the effectiveness of the self-supervised attention mechanism in enhancing BERT's generalization capabilities. The paper introduces Self-Supervised Attention (SSA) as a method to enhance BERT models' performance on smaller datasets. SSA integrates auxiliary knowledge into BERT's attention layer without requiring additional data or annotations. The SSA-Hybrid model, proposed in the paper, demonstrates improved performance compared to the base BERT model and other existing works on sentiment analysis tasks and the General Language Understanding Evaluation (GLUE) benchmark. Overall, the paper presents SSA as a promising technique for enhancing the performance of BERT models across various natural language processing tasks.

III. PROPOSED SYSTEM

The AI system designed for students incorporates a multifaceted approach to enhance the learning experience. The initial module, Document Feature Extraction, efficiently identifies and extracts key features from educational documents, such as headings, keywords, and phrases. This aids in organizing and structuring information for improved comprehension, offering students a more navigable learning resource. Named Entity Recognition (NER) [5] is a crucial component that focuses on categorizing entities like names, dates, locations, and specialized terms within the text. By accurately identifying and classifying these entities, the system enhances its ability to comprehend specific information, contributing to a more nuanced understanding of the material. Continuous improvement is facilitated through the fine tuning module, which refines the system based on user interactions and feedback. This adaptive learning process ensures that the AI becomes

progressively more attuned to user preferences and requirements, thereby optimizing its performance over time.

The Language Model (LLM) serves as a key element in facilitating comprehensive note comprehension. By enhancing the system's understanding of context, the LLM enables the AI to grasp nuanced meanings and relationships between concepts. This capability results in the generation of more coherent and contextually relevant responses during interactions [8]. A valuable tool for efficient review is the Summarization Module, which condenses lengthy passages into concise summaries while preserving essential information. This feature aids students in quickly reviewing and understanding the main points of a document, streamlining the learning process. The Quiz Generation from Note Module leverages the extracted information to create interactive quizzes based on the content of the notes [6].

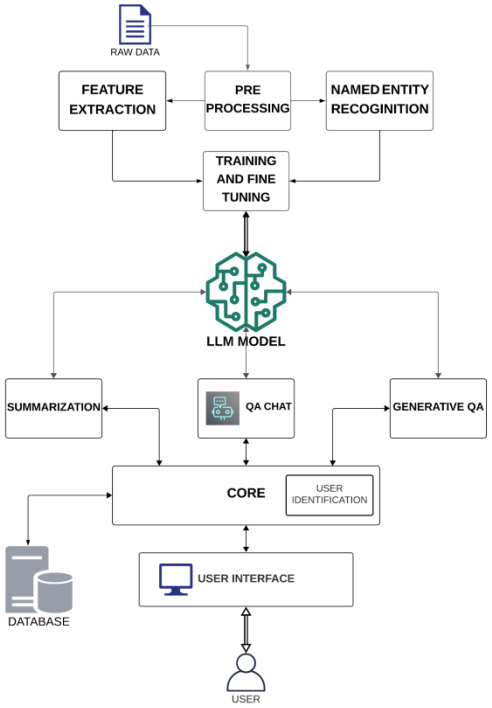


Fig 5. Illustration of proposed system

This module provides students with a valuable self-assessment tool, reinforcing key concepts and promoting active engagement with the material. To facilitate dynamic and personalized interactions, the system includes an Interactive Chat Bot with a user-friendly interface. This real-time conversational agent engages with students, offering assistance, answering questions, and providing clarifications. The user interface ensures a smooth and user-friendly experience, enabling students to effectively engage with the AI system. In essence, the holistic combination of these modules aims to optimize the learning experience by leveraging advanced language processing capabilities alongside interactive features.

IV. COMPARATIVE STUDY

The below table shows the comparison between the different models reviewed for understanding the methodologies that could be used.

TABLE I. COMPARATIVE STUDY

Methods	Advantages	Disadvantages
MRC Framework	Better performance than conventional MRC models due to Mutual Feedback	Complexity in implementation
MTA encoder for answer generation	The model achieves state-of the-art performance on the SQuAD dataset	The AG-MTA model does not outperform BERT.
T-BERTSum Model	Topic-aware text summarization model.	Lack of redundancy evaluation
Improving BERT with Self-Supervised Attention	Works exceptionally in smaller datasets	limited ability to capture intention revealed by consecutive snippets of text

V. CONCLUSION

In conclusion, the fusion of Natural Language Processing (NLP) with computer science education has unveiled a transformative landscape characterized by innovative tools and applications. Our project, leveraging state-of-the-art language models like OpenAI's GPT-3.5, has demonstrated the potential of NLP in reshaping the learning experience. Through automated summarization, dynamic question-answering, and personalized quiz generation, the system addresses the unique challenges within computer science syllabi, offering students and educators a more efficient and tailored educational journey.

While these advancements showcase the potential of NLP, challenges persist. Ethical considerations in AI-driven education and the ongoing refinement of NLP models for domain-specific tasks demand collaborative attention. The synergy of educators, researchers, and technologists is essential to navigate these challenges responsibly and foster continued innovation. As we embrace the dynamic nature of the educational landscape, there is a collective responsibility to ensure that NLP applications in computer science education contribute to inclusive, ethical, and effective learning environments.Looking forward, the future of NLP in education appears promising. Advances in language models and machine learning techniques are poised to create more adaptive and personalized learning experiences.

VI. REFERENCES

[1] Leveraging Pre-Trained Language Model for Summary Generation on Short Text: https://ieeexplore.ieee.org/document/9298823
[2] Improving BERT With Self-Supervised Attention: https://ieeexplore.ieee.org/document/9584911
[3] Multi-Layer Transformer Aggregation Encoder for Answer Generation: https://ieeexplore.ieee.org/document/9091180
[4] Machine Reading Comprehension Framework Based on Self-Training for Domain Adaptation: https://ieeexplore.ieee.org/document/9336622
[5] T-BERTSum: Topic-Aware Text Summarization Based on BERT: https://ieeexplore.ieee.org/abstract/document/9464764

[6] A Novel Document-Level Relation Extraction Method Based on BERT and Entity Information: https://ieeexplore.ieee.org/document/9098945
[7] A More Robust Model to Answer Noisy Questions in KBQA: https://ieeexplore.ieee.org/document/10058914
[8] Transformer-Based Neural Network for Answer Selection in Question Answering:https://ieeexplore.ieee.org/document/8648373
[9] Answer Category-Aware Answer Selection for Question Answering: https://ieeexplore.ieee.org/document/9245487
[10] Improving BERT-Based Text Classification With Auxiliary Sentence and Domain Knowledge: https://ieeexplore.ieee.org/document/8903313
[11] End-to-End Open-Domain Question Answering with BERTserini: https://www.sciencedirect.com/science/article/abs/pii/S0167923623001136
[12] Building a deep learning-based QA system from a CQA dataset: https://www.sciencedirect.com/science/article/abs/pii/S0167923623001136

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

Beyond Illusions: Enhancing Deepfake Detection with fine-tuned Vision Transformers

Lithiya Sara Babu[1], Jessy Willy[2], Kenza Zakeer[2], Noya Mathew[2]
[1]Assistant Professor, Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Muvattupuzha, Kerala, India
[2]Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Muvattupuzha, Kerala, India

ABSTRACT
Deepfake technology poses a significant threat to the authenticity of multimedia content, warranting robust detection mechanisms to safeguard against its misuse. This research paper conducts a comprehensive survey of deepfake detection methodologies, with a particular emphasis on the fusion of Transformer encoders and Convolutional Neural Networks (CNNs). Through an in-depth analysis of existing approaches, we identify the strengths and limitations of current techniques in the face of evolving deepfake generation methods.
This survey paper delves into the landscape of deepfake detection methodologies, with a particular focus on the integration of Transformer encoders and Convolutional Neural Networks (CNNs). Examining existing literature reveals the strengths and limitations of current approaches, underscoring the need for advancements in addressing the evolving challenges posed by deepfake generation. As a solution we propose a deepfake detection model based on a fine-tuned version of Convolutional Vision Transformer (CVT) architecture. Our approach involves optimizing learning rates and leveraging a diverse dataset that encompasses subclasses of deepfakes generated through various techniques. This work contributes to the ongoing efforts in fortifying digital content authenticity, fostering trustworthiness in multimedia platforms in the face of a rapidly evolving landscape of deepfake threats.
Keywords—Deepfake detection, CNN, Transformer Encoder, Vision Transformer, CViT

I. INTRODUCTION

Deepfake is a technique that uses deep learning algorithms to create realistic-looking fake videos or images by superimposing or manipulating existing content. It involves training a neural network on large datasets to generate convincing simulations of someone's likeness, making it appear as if they are saying or doing things they never did. Deepfakes raise concerns about misinformation, privacy, and the potential for malicious use in spreading fake content.
Deepfakes are created by training deep learning models, typically using Generative Adversarial Networks (GANs), on large datasets of a person's images or videos. These models learn to generate realistic facial

expressions, movements, and voice patterns, allowing them to convincingly replace or manipulate the original content with the desired alterations.

Current Deepfake detection techniques often involve analysing inconsistencies in facial features, unnatural facial movements, or artifacts introduced during the generation process. Some methods focus on detecting subtle discrepancies in eye blinking, lip synchronization, or facial expressions that may appear unnatural in a deepfake video. Audio-visual synchronization analysis is another approach, examining whether the audio matches the lip movements accurately.

There are three categories of existing deepfake detection models. The first approach analyses physical behaviours and psychological characteristics in videos. This includes eye blinking and head pose characteristics. The second approach involves detecting biological signals found in images. This includes features such as blood flow and GAN fingerprints. Finally, the third approach centres on identifying visual artifacts through data-driven techniques, often requiring a substantial amount of data for training. However, these methods have been criticized for their lack of generality, as they may not reliably detect multiple spoofing techniques or unseen detection methods. Additionally, there is a noted emphasis on presenting the proposed architectures, with less focus on the impact of data preprocessing on the final detection model. Recent researches have also seen the rise of novel methodologies that focuses on the identity of the subject of the deepfake rather than the composition of it such as [5].

Furthermore, recent research has highlighted the need for deepfake detection methods to address the challenge of generality, ensuring their reliability in detecting various spoofing techniques and unseen detection methods. This emphasis on generality underscores the importance of developing detection models that can effectively identify deepfakes across different settings, environments, and orientations. Additionally, there is a growing recognition of the significance of data preprocessing in enhancing the performance and robustness of deepfake detection models. These insights have spurred the development of new approaches, such as the proposed Convolutional Vision Transformer, which aims to address the limitations of existing methods by leveraging the combined learning capacity of CNNs and Transformers while emphasizing the impact of data preprocessing on the detection model.

This research paper surveys some existing deepfake techniques with emphasis on CNNs and Transformer Encoders to investigate the efficiency of various deepfake detection methods on deepfakes generated through different techniques. And from the observations made, the paper proposes a system which tries to optimize the learning rate parameter for maximum accuracy and investigate the effect of dataset on deepfake detection. Section II of the paper consists of the analysis done on of various deepfake detection methods, followed by Section III which describes the proposed system and the paper concludes with Section IV which compares the various techniques referenced in the paper.

II. LITERATURE SURVEY

A. CONVOLUTIONAL VISION TRANSFORMER

The rapid advancement in image and video alteration technologies, particularly the emergence of Deepfake, has led to hyper-realistic digital contents in the internet. Deepfake, a technique manipulating faces in videos, poses both creative and deceptive potential, being used in CGI, VR, AR, and arts while carrying risks for malicious misuse. Existing Deepfake detection methods, though are focused on tool behavior, lack generality and

robustness. Studies emphasize the weaknesses in data preprocessing and model generality. Addressing these gaps, this paper proposes a Convolutional Vision Transformer (CViT) architecture. The approach harnesses CNN and Transformer architectures, enabling holistic image feature learning and leveraging the Transformer's attention mechanism. Emphasis on comprehensive data preprocessing and training on diverse datasets aims to enhance the model's ability to detect Deepfakes across various settings and orientations. This approach is focused to develop a more generalized and robust detection framework for Deepfake videos. Fig. 1 shows the architecture of Convolutional Vision Transformers.

1) *Convolutional Neural Network (CNN):* The CNN architecture, or feature learning (FL) component, consists of 17 convolutional layers that extract low-level features from face images. It follows a VGG-like architecture without fully-connected layers, intended for feature extraction rather than classification. Also, utilizes batch normalization, ReLU activation, and max-pooling to reduce dimensionality. Consists of three consecutive convolutional operations per layer, except for the last two layers which have four operations.
2) *Vision Transformer (ViT):* In the Vision Transformer (ViT) component, the output feature maps are taken, splits them into patches, and embeds them into a linear sequence. Adds position embedding to store positional information. It then uses a Transformer encoder with Multihead Self-Attention (MSA), Multi-Layer Perceptron (MLP) blocks, and normalization layers. A MLP head performs final classification between "Fake" or "Real" face images, with Softmax applied for output.

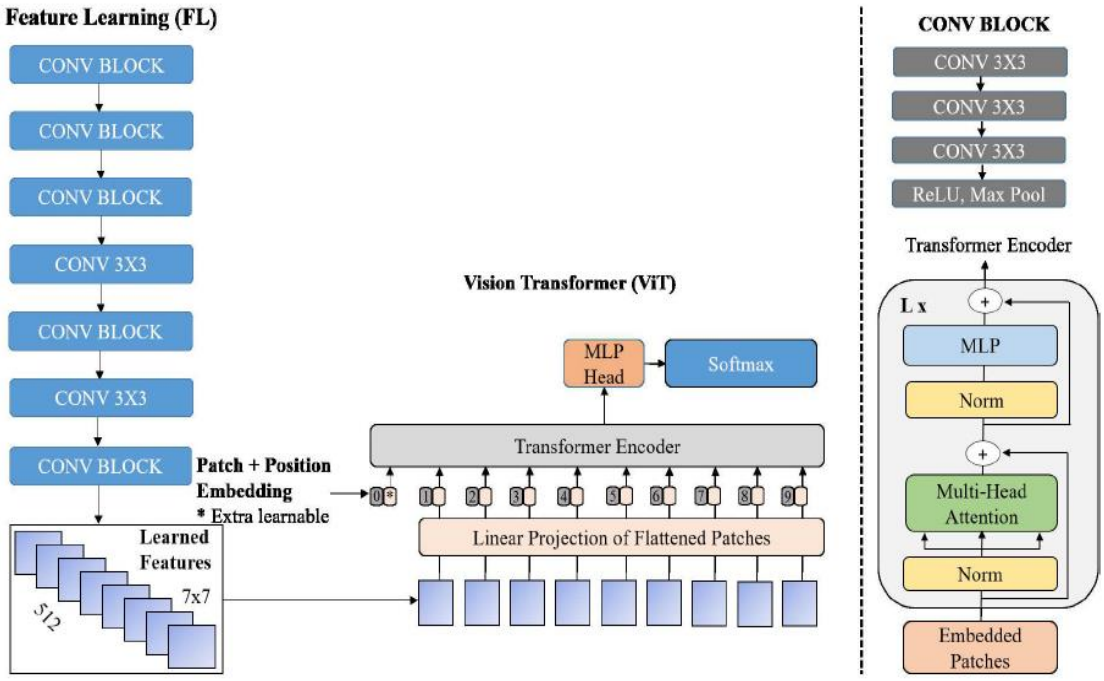


Figure 1: Convolutional Vision Transformer

B. EFFICIENTNET AND VISION TRANSFORMERS

In the sphere of computer vision, convolutional neural networks (CNNs) were used for detection of video deepfakes in videos. However, recent research has shown that Vision Transformers (ViTs) can also be effective for this task. ViTs are a type of neural network that use self-attention mechanisms to process images, allowing them to capture global context and long-range dependencies. Researchers have a new approach to video deepfake detection that combines ViTs with a convolutional EfficientNet B0. EfficientNets are a family of CNNs that have been optimized for both accuracy and efficiency, making them well-suited for resource constrained environments.

1) *Efficient ViT Model:* The architecture of the Efficient ViT consists of two key elements: a convolutional module functioning that act is a feature extractor and the other a Transformer Encoder. Inspired by the achievements of the EfficientNet, the decision is made to utilize an EfficientNet B0, the most compact variant in the EfficientNet series, as the feature extractor to process input faces. The EfficientNet generates visual features for each 7x7-pixel segment of the input face. After linear projection, each feature of each spatial location is further processed using the Vision Transformer. The binary classification score is obtained from the CLS token. Figure 2(a) provides a visual representation of the architecture. The initialization of the EfficientNet B0 feature extractor contains pre-trained weights that are then fine-tuned to optimize the performance of the final layers for a specific downstream task. The use of EfficientNet B0 convolutional network features streamlines Vision Transformer training by enhancing important low-level and local information contained in CNN features from the image.
2) *Convolutional Cross ViT:* In the context of deepfake detection, relying solely on small patches that are seen in the Efficient ViT, can turn out not to be the most optimal strategy, given the potential for artifacts to manifest both locally and globally during deepfake generation. To address this, the Convolutional Cross ViT architecture is introduced. Expanding on the foundations of both the Efficient ViT and the multiscale Transformer architecture, the Convolutional Cross ViT incorporates two separate branches: the S-branch which is dedicated for the smaller patches, and the L-branch which is designed for larger patches to encompass a broader receptive field. The outputs of visual tokens from Transformer Encoders in these two branches undergo cross attention, enabling direct interplay between the paths. Ultimately, the CLS tokens correlated to outputs of each branch help to generate separate logits. These logits are then aggregated and the final sigmoid function gives the ultimate probability scores. Convolutional Cross ViT uses two different CNN backbones—EfficientNet B0, which handles 7 x 7 image patches for the S-branch and 54 x 54 for the L-branch, and Wodajo et al.'s CNN, which handles 7 x 7 image patches for the S-branch and 64 x 64 for the L-branch. The mixed convolutional-transformer network combines the strengths of both EfficientNet and ViT, leading to better results than using either network type alone.

C. VISION TRANSFORMER WITH DISTILLATION

DeepFake is a technique to create fake visual content by changing the source person on a target video. These forged videos can be detected by identifying forged parts within or between the frames. To achieve this, the networks used are mainly composed of CNNs, taking into account spatial characteristics. However, vision transformers and use of distillation methodology have given higher accuracy in detecting deepfake content.

This paper adds a distill token to VIT along with patch embedding and CNN features as model input to create a more generalized and powerful model for deepfake detection.

1) *Vision Transformer:* The base network architecture used is Vision Transformer. The image is split into patches and these patches undergo linear projection and position embedding before being processed by Transformer encoder. The input to the encoder is processed by two approaches: patch embeddings (Zp) from ViT and CNN-derived features (Zf) before inputting them to encoder. The CNN features are obtained using a model like EfficientNet The features Zp and Zf are concatenated and subjected to global pooling making it to fixed-size representation suitable for the model input. The ViT encoder includes a multi-headed self-attention (MSA) which allows the model to process different parts of the input and multi-layer perceptron (MLP), which consists of two layers with Gaussian Error Linear Unit (GELU) non-linearity.
2) *Distillation:* Distillation method refers to the technique of transferring information from a teacher model to a smaller student model by training the student model to mimic the behavior of the teacher model, rather than directly optimizing it to predict true labels. In this paper distillation method is used along with ViT. The teacher network selected for distillation is EfficientNet, which is a state-of-the-art-model for deepfake detection in the DFDC dataset. The input representation (Zp $\oplus$ Zf) is augmented with additional tokens: a class token used during training by label and a distillation token. The final input Z0 includes these tokens along with learnable position embeddings (Epos). The tokens, both class and distillation, signify the probability that if a video is fake or not. During testing, distillation tokens are used instead of class tokens, demonstrating improved performance on the test dataset. It is to be highlighted that when using a CNN model as the teacher network, the Transformer (in this case, ViT) yielded superior results compared to other models.

D. MULTI-MODAL MULTISCALE TRANSFORMERS

The rise of sophisticated Deepfake techniques has raised concerns about the widespread manipulation of facial features in images, leading to the creation of highly realistic forged content that is challenging to visually discern. Traditional Deepfake forensics, relying on neural networks and focusing on cropped face regions, often struggles with limited receptive fields and fails to consider global pixel relationships. In response, this research introduces the Multi-modal Multi-scale Transformer (M2TR), a novel solution designed to overcome these limitations. M2TR employs a multi-scale architecture that utilizes transformers for effective modeling of pixel relationships. Unlike conventional vision transformers, M2TR operates on a two-stream architecture, incorporating both RGB domain and frequency information to enhance Deepfake detection. The model includes a cross-modality fusion block for efficient information integration and extends beyond binary classification by predicting manipulated regions in a multitask manner, addressing overfitting concerns. Moreover, the study presents the Swapping and Reenactment DeepFake (SR-DF) dataset, offering a substantial and well-crafted collection of data suitable for advancing the development of Deepfake detection techniques. The authors suggest a systematic assessment framework to thoroughly evaluate the dataset's quality, thereby offering valuable contributions and advancements to the realm of Deepfake detection.

1) *Multi-modal Multi-scale Transformer(M2TR):* The Multi-modal Multi-scale Transformer (M2TR) is an approach for detecting Deepfakes, intended to address the security risks associated with manipulated content. It consists of three main elements: stacked multi-scale transformers, frequency filters, and cross-modality fusion blocks. M2TR employs a multi-scale transformer to identify forgery patterns across various patch sizes, thereby improving its ability to detect manipulation.
2) *SR-DF:* The SR-DF dataset is crucial in evaluating how well the Multi-modal Multi-scale Transformer (M2TR) performs in detecting deepfakes. This extensive dataset, meticulously crafted for deepfake forensics, incorporates various synthesis techniques like face swapping and facial reenactment. It prioritizes advanced post-processing to ensure images exhibit high quality and diverse characteristics. During the evaluation phase, SR-DF surpasses existing deepfake datasets, excelling in both image quality and diversity. This establishes SR-DF as an asset for advancing research in deepfake detection. The comprehensive assessment, combined with the unique capabilities of M2TR, greatly advances the field's comprehension and abilities in tackling the challenges presented by deepfake threats.

III. PROPOSED SYSTEM

A. Architecture

Our research focuses on enhancing Deepfake detection in videos by leveraging a Convolutional Vision Transformer (CViT) model, fine-tuned using a modified dataset, FaceForensic++. Our proposed model uses a base CViT architecture that has two main components: the Feature Learning (FL) component and the Vision Transformer (ViT) component. The Feature Learning component employs 17 convolutional layers inspired by the VGG architecture, extracting low-level features from face images. This is followed by a ViT component which further contains two blocks: Multi-Head Self-Attention (MSA) and Multi-Layer Perceptron (MLP) blocks. This component processes the feature map produced by the FL component and final classification is done. The FaceForensics++ dataset and CelebDF dataset were used to create train, test, and validation datasets for our finetuned model, this constituted the preprocessing component. Two subcomponents contribute to this process are Face Extraction Component and Data Augmentation Component. Fig. 2 shows the system architecture.

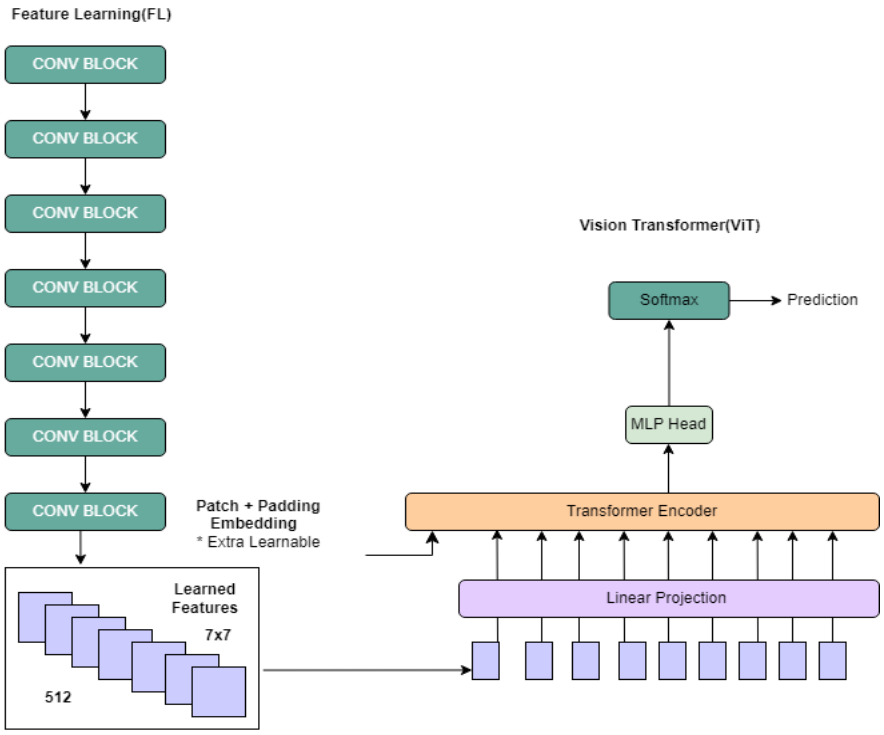


Figure 2: Architecture of Convolutional Vision Transformer

For extracting faces from videos to prepare the dataset, and for further validation and usage of the model a Face Extraction Component is used. This component ensures a standardized input format of 224 x 224 RGB. The Data Augmentation Component diversify the dataset and improve model generalization, data augmentation techniques are applied during training. The dataset transition from the original to FaceForensic++ is carried introduces novel challenges and variations, necessitating adjustments to the model. The FL component is fine-tuned to adapt to the unique features present in the FaceForensic++ dataset. The learning rate is optimized during fine-tuning. Systematic exploration of different learning rate values is conducted to strike a balance between convergence speed and avoiding overshooting.

B. Implementation

Our implementation of the proposed Deepfake detection model involves the following key steps, building upon the architecture detailed in III.A:

Video processing libraries and facial detection algorithms are used to extract face images from the FaceForensic++ dataset. Each face image is resized to a standardized 224 x 224 RGB format, aligning with the input requirements of our CViT model. During the training phase, data augmentation techniques are applied that include techniques like random rotations, flips, and changes in brightness, enriching the dataset and promoting model generalization. The ViT component follows the architecture described in III.A. The ViT encoder, has two blocks: Multi-Head Self-Attention (MSA) and Multi-Layer Perceptron (MLP) blocks. These blocks process the feature map generated by the Feature Learning component. The ViT model undergoes end-to-end training, incorporating position embeddings and passing through the Transformer for processing. Fine-tuning of the model involves systematically adjusting the learning rate, an essential hyperparameter, to attain optimal convergence. Fine-tuning is performed on the FL component to extract features specific to the characteristics of the FaceForensic++ dataset.

The model is trained on the training set, with validation performed at regular intervals to monitor progress and prevent overfitting. The testing component evaluates the model's performance on the unseen data from the FaceForensic++ dataset.

During testing, the model's predictions are compared against ground truth labels from the FaceForensic++ dataset. We compute standard evaluation metrics, such as precision, recall, F1-score, and accuracy, to offer a thorough evaluation of the model's capability in identifying Deepfake videos.

IV. COMPARATIVE STUDY

From the survey conducted about various deepfake detection and generation techniques, the conclusion arrived at was that the combination of CNN and Transformer is highly effective technique for generalized deepfake detection. Since CNN models are generally highly dataset dependent [6] and identity driven deepfake detection might prove to be computationally more complex as in [5], the combination of CNN as well as Transformer Encoder (CViT) [1] offers a more affordable and generalized model. The proposed system aims to fine-tune the existing CViT model by optimizing the learning rate to provide a more robust detection. Table I shows the comparison between different models reviewed for understanding the methodologies that could be used.

Table 1: Comparative study

Methods	Advantages	Disadvantages
Convolutional Vision Transformer	Novel Approach MFM Activation Function Robust Detection	Limited Insight into False Positives/Negatives
EfficientNet and Vision Transformers	Computationally light Efficient with respect to no. of parameters used.	Less generalizable on different techniques
Vision Transformer with distillation.	More robust, lower average loss and better f1 score.	No much improvement in accuracy compared to existing state-of-art model.
Multi-Modal Multi-scale Transformers.	High accuracy, Multi-modal architecture High-quality Deepfake dataset.	Limited availability of code and models Limited datasets
Identity-Driven DeepFake Detection	High accuracy and generalization. Focuses on identity	Computational complexity Dataset limitations
Deep Ensemble Based Learning Technique	High accuracy Real time detection	Computational resources Dataset dependency

V. CONCLUSION

In conclusion, this research paper has provided a comprehensive survey of deepfake detection techniques, specifically focusing on the usage of Transformer encoders and Convolutional Neural Networks (CNNs) and the

combination of both techniques. By reviewing existing literature, we identified the strengths and limitations of current methods in addressing the evolving challenges posed by deepfake generation.

Building upon this survey, our proposed model introduces a novel approach to enhance deepfake detection. Leveraging a fine-tuned Convolutional Vision Transformer (CVT) architecture, we optimize the learning rate to achieve improved convergence and model performance. Additionally, the utilization of a diverse dataset, containing subclasses of deepfakes generated through various techniques, allows our model to generalize more effectively across a broader range of deepfake scenarios.

The fine-tuning strategy, coupled with dataset diversification, will contribute to the model's ability to discern subtle nuances in deepfake creation techniques, thus providing a valuable advancement in mitigating the challenges posed by increasingly sophisticated deepfake generation methods.

As deepfake technology continues to evolve, we envision that the proposed model will serve as a valuable asset in the broader landscape of combating misinformation and ensuring the trustworthiness of multimedia content in the digital age. Further research avenues may explore the integration of additional modalities and continuous adaptation to emerging deepfake generation techniques to bolster the resilience of detection systems.

VI. REFERENCES

[1] Wodajo, D., & Atnafu, S. (2021). Deepfake Video Detection Using Convolutional Vision Transformer. ArXiv, abs/2102.11126.

[2] Coccomini, D.A., Messina, N., Gennaro, C., & Falchi, F. (2021). Combining EfficientNet and Vision Transformers for Video Deepfake Detection. ArXiv, abs/2107.02612

[3] Heo, Y., Choi, Y.J., Lee, Y., & Kim, B. (2021). Deepfake Detection Scheme Based on Vision Transformer and Distillation. ArXiv, abs/2104.01353.

[4] Wang, J., Wu, Z., Chen, J., & Jiang, Y. (2021). M2TR: Multi-modal Multi-scale Transformers for Deepfake Detection. Proceedings of the 2022 International Conference on Multimedia Retrieval.

[5] Dong, X., Bao, J., Chen, D., Zhang, W., Yu, N., Chen, D., Wen, F., & Guo, B. (2020). Identity-Driven DeepFake Detection. ArXiv, abs/2012.03930.

[6] Rana, M.S., & Sung, A.H. (2020). DeepfakeStack: A Deep Ensemble-based Learning Technique for Deepfake Detection. 2020 7th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2020 6th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom), 70-75.

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

Renal Vista : A Machine learning approach based on SVM and CNN for Prediction and Detection of Kidney Diseases

Elizabeth Anns[1], Athul Saji[2], Praveen Prasad[2], Rajul Racy[2]
[1]Assistant Professor, Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Ernakulam, Kerala, India
[2]Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Ernakulam, Kerala, India

ABSTRACT
Kidneys are vital organs that filter out fluid and waste from your blood. Late diagnosis of kidney diseases can lead to increased healthcare costs and premature mortality. This innovative kidney healthcare system addresses chronic kidney disease by predictive model using SVM. This system also includes identification of kidney conditions through CT scan analysis, streamlining kidney donation matching, and offering a nephrologist directory for consultations. With a focus on promoting kidney health awareness and fostering a supportive community, this system aims to empower individuals with information, resources, and a holistic solution for optimal kidney health.

Keywords: Chronic Kidney Disease(CKD), Support Vector Machine(SVM), Prediction, Convolutional Neural Networks(CNN), Machine Learning (ML), Deep Learning

I. INTRODUCTION

Kidneys play a critical role in filtering waste from the blood, balancing electrolytes, and regulating blood pressure. However, kidney disease can be difficult to detect early due to its subtle onset, often leading to delayed recognition of symptoms until later stages. This underscores the importance of improving diagnostic methods to identify the disease sooner, especially considering the risk factors such as diabetes, hypertension, and family history.

The increasing global prevalence of kidney diseases highlights the urgent need for innovative detection and management strategies. This system aims to revolutionize kidney healthcare by integrating advanced technologies.The proposed approach integrates machine learning algorithms and advanced image analysis techniques to predict, diagnose kidney disease, and match kidneys for transplantation. Utilizing Support Vector Machines (SVM) for predictive analysis and Convolutional Neural Networks (CNN) for image-based diagnosis, this system aims to improve patient outcomes and provide a comprehensive framework for addressing Chronic Kidney Disease (CKD) challenges.

In the first study [1], machine learning for CKD diagnosis is explored, focusing on Support Vector Machine (SVM). SVM demonstrates high accuracy of 99.1%, highlighting its potential for early CKD detection. The second study [2] further investigates SVM, emphasizing its superiority among machine learning algorithms. Recommendations for dataset refinement and ensemble models are made, enhancing SVM's effectiveness in optimizing CKD diagnosis.

The third study [3] introduces a specialized CNN model for distinguishing kidney disorders from CT scans. Despite recognizing limitations, the study underscores CNNs' potential in advancing medical diagnostics. The fourth study [4] introduces the Adaptive Hybridized Deep Convolutional Neural Network (AHD CNN) for CKD diagnosis, showcasing CNNs' ability to address challenges posed by renal cancer comprehensively.

The fifth study [5] addresses CKD challenges, with a focus on early detection to prevent adverse outcomes. Using machine learning, it employs models like Random Forest (RF), Gradient Boosting (GB), XGBoost (XGB), Logistic Regression (LR), and Support Vector Machine (SVM) on a dataset from Apollo Hospitals, Tamil Nadu, India, containing 24 attributes. The study demonstrates machine learning's potential to enhance diagnosis, reduce costs, and enable early intervention.

II. LITERATURE SURVEY

A. Comparison Between SVM, KNN And Decision Tree Algorithms

[1]Exploring machine learning techniques for diagnosing kidney disease, the first study employs a robust set of four key algorithms: Support Vector Machine (SVM), Decision Tree, Random Forest, and K-Nearest Neighbour (KNN). SVM, a cornerstone of the analysis, stands out for its prowess in classifying data. It achieves this by establishing a hyperplane with maximum margin, allowing it to handle both linear and non-linear scenarios effectively. The study places particular emphasis on the intricate process of feature selection, highlighting its crucial role in optimising model efficiency and interpretability. Remarkably, this meticulous approach results in a reported accuracy rate of 99.1% biological image recognition, the suggested system remains broadly applicable and functional across different datasets.

However, as with any scientific endeavour, the study candidly acknowledges certain limitations. Notably, the dataset employed consists of a relatively small sample—only 400 occurrences. This raises valid concerns about the model's generalisation capability to a broader population. Moreover, the absence of external validation adds complexity to the assessment of the model's robustness when exposed to novel data. The architectural framework, elegantly depicted in Figure 1, provides a visual representation of the system's inner workings

B. Comparison Between SVM, KNN, Naive Bayes and Linear Regression Algorithms

[2]Transitioning to a broader healthcare perspective, the second study delves into the intricate landscape of CKD as a global health concern. To address this pervasive issue, the study proposes the incorporation of machine learning (ML) techniques for early detection. The chosen quartet of ML algorithms includes Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Naive Bayes, and Linear Regression. The methodology adopted encompasses a systematic and thorough process, covering dataset collection, preprocessing steps, data splitting, ML model selection, and comprehensive outcome analysis.

Within this framework, SVM emerges as a standout model, gaining prominence for its exceptional performance compared to its counterparts—Naive Bayes, KNN, and Linear Regression. The study not only positions SVM as

a superior choice for CKD detection but also offers forward-thinking recommendations. Future enhancements are suggested, with a specific focus on refining the dataset and exploring ensemble models.The main objective is to improve prediction performance, which supports the study's central hypothesis that SVM can significantly improve the effectiveness and precision of CKD diagnosis in the context of ML algorithms

C. Convolutional Neural Network

[3]The first study addresses the escalating global incidence of kidney illness, emphasising the pressing need for efficient detection techniques. In a groundbreaking move, the study introduces a customised convolutional neural network (CNN) model. designed explicitly for distinguishing kidney disorders from Computed Tomography Scan. This pioneering approach involves a sophisticated combination of deep learning principles and CNNs, particularly harnessing the power of ResNet50, a 50-layer convolutional neural network, for image categorization.

The suggested methodology takes a novel step by dividing kidney CT scans into four distinct groups—cyst, normal, stone, and tumour This classification is achieved through a critical preprocessing procedure known as "watershed segmentation." The segmentation process plays a pivotal role in isolating the image's region of interest, facilitating more accurate categorization. Despite achieving a notable accuracy rate of 98.66% on the test set, the study candidly acknowledges several drawbacks. Issues such as over-segmentation and noise sensitivity are highlighted. Furthermore, a crucial question is raised about the model's generalizability to a variety of clinical contexts, given the exclusive reliance on a single Kaggle dataset for both testing and training. However, the study's findings highlight the revolutionary potential of machine learning in increasing early sickness prediction, a huge step in medical diagnostics.

D. Adaptive Hybridized Deep Convolutional Neural Network

[4]The second study introduces a cutting-edge model, the AHD CNN, specifically designed for the diagnosis of kidney disease. This adaptive system, tailored for superior object recognition in kidney disease images, incorporates an array of advanced techniques. These include convolutional layers with spatial max pooling, group normalisation, ReLU gating, and linear convolution. The research explores CNN's pivotal role in dynamically adjusting the Convolutional Neural Network (CNN) during training, enabling the effective utilisation of unlabeled data.

The study's primary objective is ambitious—to map kidney cell activity representations from MRI signals to hybridised functionality derived from CNNs. This intricate mapping process is particularly crucial for consistently forecasting renal cell ratings, especially in contrast-enhanced computed tomography (CECT) images for tumour analysis. The study demonstrates effective tumour analysis, showcasing the comprehensive nature of the research. The mechanics of CNNs are thoroughly explored, emphasising their significance in signal processing tasks. Various techniques, including atrous convolution, deep residual learning, and skip connections, are meticulously examined to enhance model accuracy. The segmentation pipeline incorporates advanced preprocessing methods such as histogram equalisation, nonparametric bias correction, and autoencoders, overcoming challenges like low contrast and artifacts.

Moreover, the study ventures into the realm of predicting patients' survival time directly from pathological images for renal cancer. This involves a multifaceted approach, encompassing 2D and 3D CT image analysis, feature extraction using CNNs, feature fusion, and classifier feeding. Techniques like support vector

classification, multi-crop pooling layers, and thresholding contribute to the successful detection and analysis of tumours. This comprehensive research not only delves into the intricacies of CKD diagnosis but also showcases the potential of CNNs to offer a holistic and accurate solution to the challenges posed by renal cancer

III. PROPOSED SYSTEM

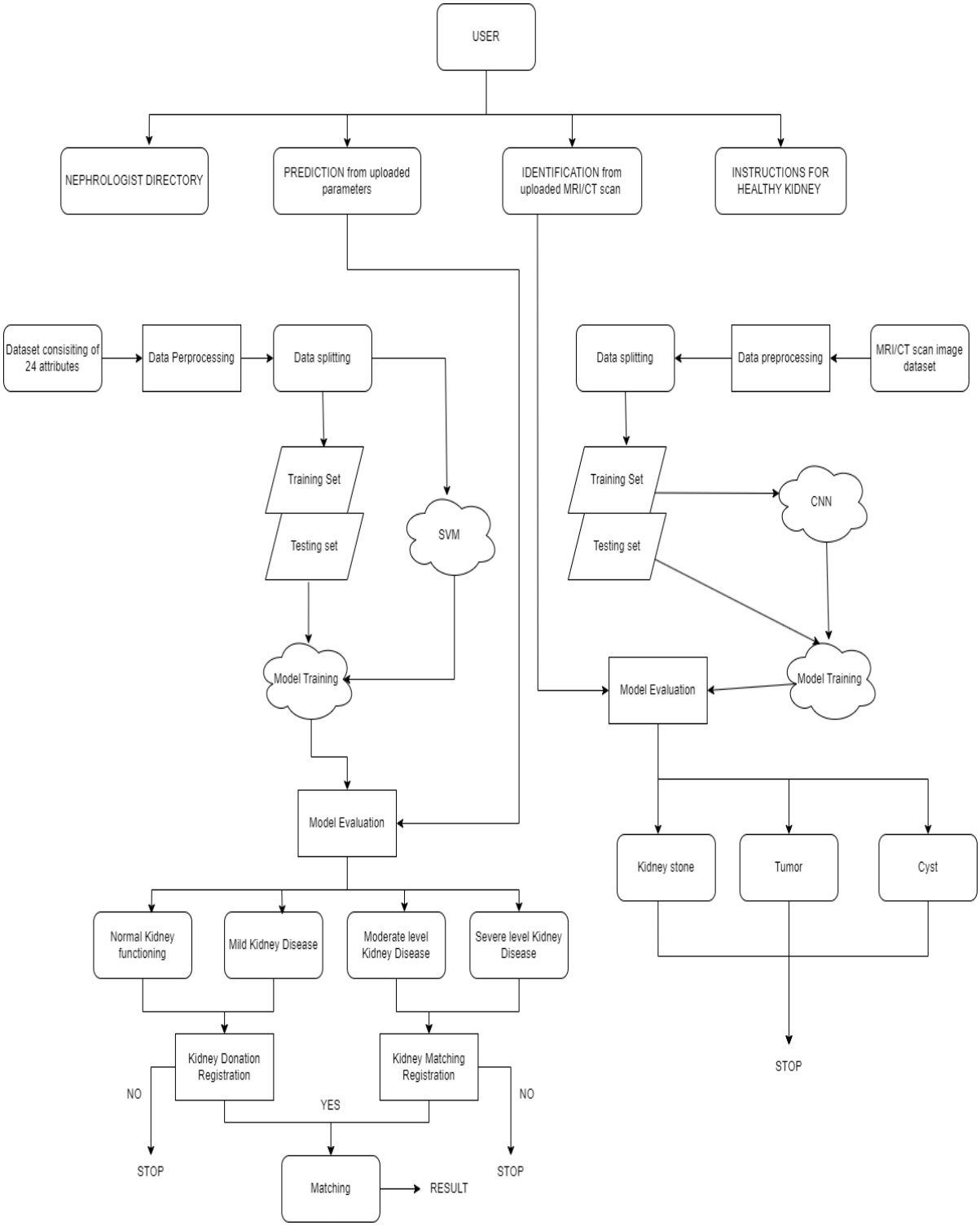


Figure 1: Architecture

A. Architecture

The system operates on a robust and dynamic architecture, seamlessly integrating its four key features. Upon user input of medical parameters, a sophisticated Support Vector Machine (SVM) model processes the data to predict the likelihood of future kidney diseases. Simultaneously, advanced image processing algorithms meticulously analyze CT scan images to identify kidney stones, cysts, and tumors. The system's architecture includes a centralized repository housing comprehensive information and contact details for nephrologists. The innovative preliminary donor matching feature utilizes various parameters to expedite the pairing of potential donors with recipients. This architecture ensures a fluid flow of information, promoting collaboration and active user participation while providing a holistic and proactive approach to kidney healthcare.

B. Implementation

Our proposed system is implemented using a sophisticated and modular approach, ensuring seamless integration of its diverse functionalities. The predictive analysis component utilizes machine learning algorithms, particularly the Support Vector Machine (SVM) model, trained on a diverse dataset to enhance accuracy in predicting kidney diseases based on input medical parameters.

For CT scan image analysis, cutting-edge image processing algorithms identify and categorize kidney stones, cysts, and tumors with high precision, facilitating prompt and targeted medical interventions. The system's repository, hosting information on nephrologists, is designed with a scalable and secure architecture for efficient storage and retrieval of data. This directory ensures quick access to vital contacts, fostering effective communication between patients and specialists.

The preliminary donor matching feature integrates user and donor information from the system's repository, expediting the matching process for potential kidney transplantations. Overall, the implementation of the Kidney Healthcare System harmoniously integrates machine learning, image processing, and robust data management, providing an effective, user-friendly, and collaborative solution for proactive kidney care.

TABLE I COMPARATIVE STUDY OF LITERATURE SURVEY

Prediction Models	Accuracy	Advantages	Disadvantages
Comparison Between SVM, KNN And Decision Tree Algorithms	99.1%	Early detection of kidney disease	Limited dataset of 400 occurrence, Lack of real-world validation
Comparison Between SVM, KNN, Naive Bayes and Linear Regression Algorithms	93.4%	Accurate classification	Does not address the imbalanced nature of the datasets, which may lead to biased or unstable results.
Convolutional Neural Network	98.66%	High Accuracy	Performance of the proposed model is not analysed with the exiting the model
Adaptive Hybridized Deep Convolutional Neural Network	90%	Save time and reduce the risk of human error	Requires a large dataset

IV. CONCLUSION

In conclusion, this comprehensive survey meticulously explores various models for the detection of chronic kidney disease in healthcare, presenting a diverse array of methodologies and thoroughly evaluating their respective strengths and limitations. The findings highlight the pivotal role of technology in addressing the challenges associated with chronic kidney disease detection. This survey underscores the potential for synergies between medical science and technological advancements in healthcare.

V. REFERENCES

[1] C. P. Kashyap, G. S. Dayakar Reddy and M. Balamurugan, "Prediction of Chronic Disease in Kidneys Using Machine Learning Classifiers," 2022 1st International Conference on Computational Science and Technology (ICCST), Chennai, India, 2022, pp. 562-567, doi: 10.1109/ICCST55948.2022.10040329.

[2] M. Kaur and G. Kaur, "Comparative Analysis of Machine Learning Algorithms for Early Detection of Chronic Kidney Disease," 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2023, pp. 1754–1759, doi: 10.1109/ICESC57686.2023.10193695.

[3] M. S. Hossain, S. M. Nazmul Hassan, M. Al-Amin, M. N. Rahaman, R. Hossain, and M. I. Hossain, "Kidney Disease Detection from CT Images Using a Customized CNN Model and Deep Learning," 2023 International Conference on Advances in Intelligent Computing and Applications (AICAPS), Kochi, India, 2023, pp. 1-6, doi: 10.1109/AICAPS57044.2023.10074314.

[4] G. Chen et al., "Prediction of Chronic Kidney Disease Using Adaptive Hybridized Deep Convolutional Neural Network on the Internet of Medical Things Platform," in IEEE Access, vol. 8, pp. 100497-100508, 2020, doi: 10.1109/ACCESS.2020.2995310..

[5] M. Rashed-Al-Mahfuz, A. Haque, A. Azad, S. A. Alyami, J. M. W. Quinn and M. A. Moni, "Clinically Applicable Machine Learning Approaches to Identify Attributes of Chronic Kidney Disease (CKD) for Use in Low-Cost Diagnostic Screening," in IEEE Journal of Translational Engineering in Health and Medicine, vol. 9, pp. 1-11, 2021, Art no. 4900511, doi: 10.1109/JTEHM.2021.3073629.

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

Human Activity Recognition Model in Medical Sector

Swathy Venugopal[1], Aswathy Murali[2], Gishna Biju[2], Norah Jomon[2], Renjima Reji[5]
[1]Assistant Professor, Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology Vazhakulam Muvattupuzha, Kerala, India
[2]Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology Vazhakulam Muvattupuzha, Kerala, India

ABSTRACT
This paper proposes a framework leveraging MediaPipe Holistic technology for robust human action detection under varying conditions and viewing angles. MediaPipe Holistic provides pose, face, and hand landmark detection models, yielding 501 landmarks from real-time device feeds. To ensure consistency and reliability in the study, both models are trained using identical datasets derived from wearable sensors, sourced from a publicly available repository.[2] The evaluation includes a comparison of their accuracy and confusion matrix to assess true and false positives. Furthermore, the paper explores different aspects and domains where these models can be applied either individually or collaboratively within the broader context of Human Activity Recognition utilizing image data. The proposed framework showcases the potential of integrating MediaPipe Holistic technology with machine learning algorithms for effective HAR systems, with implications for domains like security, healthcare, and IoT.

Keywords: Long-Short term memory, Machine Learning, Gate Recurrent Unit, Convolutional Neural Network, MediaPipe, Human pose

I. INTRODUCTION

Human Activity Recognition (HAR) has emerged as a critical area of research within computer science and machine learning, finding applications in diverse fields such as surveillance, human-robot interaction, and healthcare. The ability to accurately identify human activities is essential for developing intelligent systems capable of assisting humans in various aspects of daily life. While existing HAR methods show promising results, they often demand significant computational resources due to the complexity of the task.

MediaPipe presents a novel approach to HAR, offering a machine-learning framework tailored for real-time video inputs, dealing effectively with time series data. Its versatility extends to applications such as hand tracking, face detection, pose estimation, object detection, and augmented reality. Complementing Media Pipe is OpenCV, a robust and open-source library for computer vision and machine learning, equipped with numerous algorithms and tools for image processing and objects detection.

In the proposed HAR system,[4]MediaPipe analyzes preprocessed video inputs with the aid of OpenCV, harnessing the strengths of both technologies to detect key features in human bodies and enhance the

robustness and accuracy of real-time human activity recognition. Deep learning techniques, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have demonstrated significant promise in HAR by automatically extracting features from sensor or video frame data, thus eliminating the need for manual feature engineering.

Recent research efforts in HAR focus on developing efficient, scalable, and robust systems. One promising approach involves the utilization of lightweight neural networks that can operate on low-power devices, facilitating deployment in various environments. As the field of HAR continues to evolve rapidly, ongoing research endeavours aim to introduce novel methods and techniques, continually refining and expanding the capabilities of human activity recognition systems.

II. LITERATURE SURVEY

A. ANALYSIS OF ANTENNA FOR ADAPTIVE ELIMINATION

The setup has many transmitters and receivers, using data sets like StanWiFi and "MultiEnvironment," each with one transmitter and three receiver antennas. When people move around, it can change Wi-Fi signals going from the transmitter to the receiver, affecting how well the antennas pick up the signals depending on things like which way the person moves and how high up they are. Different activities affect the antennas differently, so past studies have looked into things like picking specific parts of the Wi-Fi signal and combining them. Even though picking specific parts of the signal might not make a big difference for antennas that aren't very sensitive, it shows that antennas react differently to the same things people do. While many Human Activity Recognition (HAR) systems use accelerometers and gyroscopes for motion data, antenna analysis focuses on characteristics like amplitude, frequency, and patterns associated with activities.[1]Adaptive techniques filter noise from sensor data, crucial for accurate activity recognition due to potential interference. Adaptive algorithms extract relevant features, encompassing patterns, magnitudes, or time intervals linked to diverse activities. Machine learning tools like decision trees or neural networks learn from customized sensor data to recognize particular activities better. By improving pre-processing through adaptive elimination, we enhance the performance of the classifier.

B. STATE OF THE CHANNEL INFORMATION (SCI)

State of the Channel State Information (CSI) tells us about how good the connection is between devices, like phones or computers, when people move around.[3]It considers things like how far apart the devices are, how signals bounce around, and how strong the signals are. CSI is useful for recognizing what people are doing, understanding gestures, and figuring out where someone is because it's sensitive to changes in the environment. Nowadays, our wireless technology uses something called MIMO, which uses many antennas to send and receive signals. Each pair of antennas creates a way to talk using different methods, and one common method is called OFDM. OFDM lets us send lots of information at once on different frequencies, and CSI helps us keep track of all these frequencies.

C. CONVOLUTIONAL NEURAL NETWORK

The new deep CNN-LSTM model with self-attention is the first step in recognizing what people are doing. This model uses a CNN with 16 filters and each filter looks at a small piece of the data to figure out important

features. The rectified linear units act as non-linear activation function for the convolutional layers, it aiming to capture hidden features in input samples.[2] Convolutional Neural Networks (CNNs) have demonstrated high effectiveness in Human Activity Recognition. In the HAR context, CNNs excel in learning hierarchical features from time-series sensor data like accelerometer and gyroscope readings. The convolutional layers of the network are capable of capturing local patterns and variations in input sequences, enabling the model to discern intricate details related to diverse human activities. Moreover, CNNs can adaptively extract spatial and temporal features, essential for recognizing complex and dynamic movement patterns. The application of 1D convolutions is common in HAR, facilitating the efficient processing of sequential sensor data. The integration of CNNs enhances the robustness and accuracy of HAR systems, contributing to advancements in health monitoring, sports analysis, and context-aware computing.

D. LSTM NETWORK STRUCTURE

The suggested deep CNN-LSTM model with self-attention is designed for recognizing human activities. It uses a Long Short-Term Memory (LSTM) network, which is a type of Recurrent Neural Network. The LSTM network is important for understanding time-related information from the output of the convolutional layer. It has shown great performance with different time-related strategies.[4] LSTMs are well-known for their capability to understand long-term connections and patterns over time, making them very effective for Human Activity Recognition (HAR).Given that human activities are often characterized by specific movement sequences over time, LSTMs excel at learning and retaining these temporal dependencies. Their automatic extraction of relevant features from sequential sensor data, encompassing temporal patterns and dynamic aspects, proves essential for accurately discerning between different human activities in HAR applications.

E. DEEP ConvLSTM HYBRID MODEL

The Deep ConvLSTM hybrid model represents a significant advancement in activity recognition by seamlessly integrating Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. This fusion leverages the strengths of both architectures, positioning the model as a cutting-edge paradigm in human activity recognition. The model's architecture features four convolutional layers strategically placed atop raw sensor channels, enabling the capture of intricate spatial features.[5] This spatial sensitivity is crucial for understanding the nuances of human movements. Following the convolutional layers, two LSTM layers are employed to capture long-term dependencies within sequential sensor data, making them instrumental in recognizing activities that unfold over time. The synergy between CNNs and LSTMs enables the model to excel in both local pattern recognition and abstract modelling, showcasing a harmonious integration of spatial and temporal information. This comprehensive approach positions the Deep ConvLSTM model as a formidable contender in human activity recognition, promising enhanced accuracy and robustness. Exploring the setups of both the CNN and Deep ConvLSTM models provides insight into the unique strengths and roles of each component, highlighting the sophisticated techniques employed to achieve superior performance and marking a significant milestone in the evolution of activity recognition methodologies.

F. IDENTITY BASED TRANSFER SYSTEM

The identity attribute's too compressed state is known as identity (ID) in index form.Its magnitude and continuity were lost due to overcompression, making it nearly hard for neural networks to interpret and draw

conclusions from. It entails tracking and preserving the relationship between people during multiple video frames. People are given unique identifiers by it, which guarantees consistent recognition even in the event of occlusions or position alterations. With the help of this method, identifying information may be smoothly transferred from one frame to the next, enabling precise tracking of people throughout the video sequence.[4] Classifier structure and training approach: Attention processes, identity embedding, feature extraction, and temporal context integration are all part of the classifier structure. For best results, the training technique makes use of transfer learning, margin losses, and identity loss functions. Sequential training guarantees data and frame consistency. Consistency across frames is guaranteed by sequential training, and resilience to a variety of circumstances is improved via data augmentation. Techniques for regularization stop over fitting. By taking a comprehensive approach, the model is guaranteed to capture unique traits that enable precise recognition of human activity and transfer identities between successive frames in dynamic contexts.

III. PROPOSED SYSTEM

A. Architecture

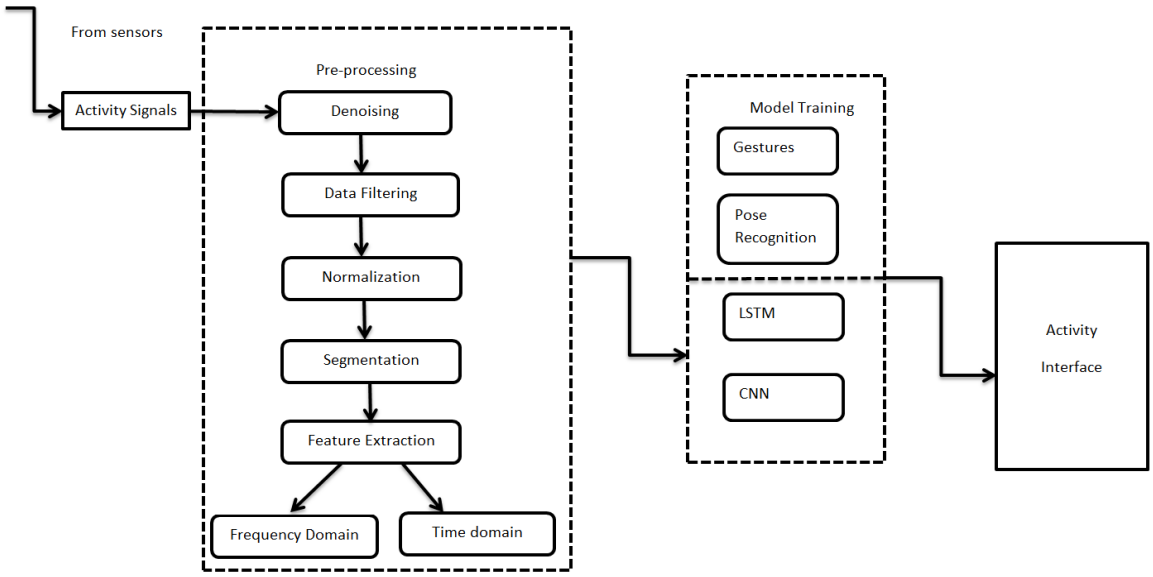


Figure 1: Architecture diagram of HAR using mediapipe

B. Implementation

The modules of the proposed system include:

1. **Data Collection:** We collect data from various sources such as wearable devices, smartphones, or CCTV cameras. The data collected could include motion data, video data, or other contextual data such as GPS location or environmental factors. The data collection module should ensure that the data is collected in a reliable and consistent manner, with minimal loss or corruption.

2. **Data Pre-processing:** Here, we clean and prepare the raw data collected in the data collection module. Pre-processing techniques could include noise reduction, data filtering, and segmentation. Noise

reduction techniques such as denoising are used to remove irrelevant information from the data. Data filtering techniques such as smoothing are used to remove high-frequency noise from the data. Segmentation techniques such as sliding window are used to divide the data into smaller segments or windows.

3. **Feature Extraction:** We extract meaningful features from the pre-processed data. Features could include statistical measures, frequency-domain measures, or time-domain measures. Statistical measures such as mean, variance, and skewness could be used to describe the distribution of the data. Frequency-domain measures such as Fourier transform or wavelet transform could be used to analyse the frequency content of the data. Time-domain measures such as autocorrelation or cross-correlation could be used to analyse the temporal dependencies in the data. The feature extraction module should ensure that the features extracted are relevant and discriminative for the activity recognition task.

4. **Model Training:** This module is responsible for training machine learning models to classify the extracted features into different activity categories. Techniques could include supervised learning, unsupervised learning, or deep learning. Supervised learning techniques such as support vector machines or decision trees could be used to learn a mapping from features to activities. Unsupervised learning techniques such as clustering could be used to discover patterns or structures in the data. Deep learning techniques such as convolutional neural networks or recurrent neural networks could be used to learn complex representations of the data.

5. **Activity Interface:** This module is responsible for presenting the results of the activity recognition to the user. This could be done through a user interface (UI) that displays the recognized activities and other relevant information. The activity interface module should ensure that the UI is intuitive and user-friendly, with clear and concise feedback to the user.

The below figure (Figure: 2) is a hierarchical organization of body parts, which may represent a diagram or image of a humanoid figure. The structure consists of a collection of connected elements, such as a torso, limbs, and a head, forming a coherent whole. The connections between the parts indicate a spatial relationship, with parent and child elements defining a parent-child hierarchy.

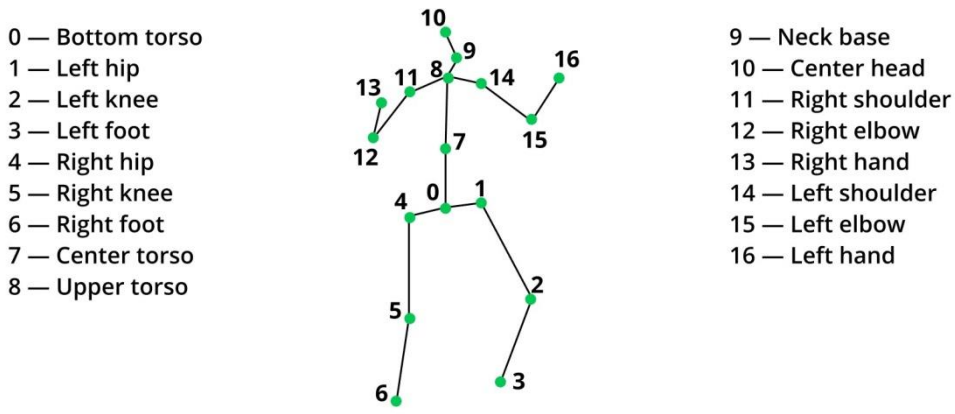


Figure 2: 3D Human Pose Estimation Model

IV. COMPARATIVE STUDY

SL.NO	PAPER TITLE	ADVANTAGES	DISADVANTAGES
1.	DEEP CNN-LSTM WITH SELF-ATTENTION MODEL FOR HUMAN ACTIVITY RECOGONITION USING WEARABLE SENSOR	• High Accuracy • Innovative Model Design	• Validation and Testing Transparency • Real-world Implementation Challenges
2.	SENSCAPSNET: DEEP NEURAL NETWORK FOR NON OBTRUSIVE SENSING BASED HUMAN ACTIVITY RECOGNITION	• Accurate Activity Recognition • Real-Time Recognition	• Data Dependency • Computational demand
3.	DEVICE FREE HUMAN ACTIVITY RECOGNITION WITH IDENTITY BAED TRANSFER MECHANISM	• Extensive experiments • High recognition rates	• Limited practical application • Lack of comparison
4.	EFFICIENT WI-FI-BASED HUMAN ACTIVITY RECOGNITION USING ADAPTIVE ANTENNA ELIMINATION	• Non-Intrusive • Monitoring Low-Cost Solution	• Limited Accuracy • Sensitivity to Environment Changes

V. CONCLUSION

Human Activity Recognition (HAR) is a continually evolving field, driven by advancements in tools and technologies that enable faster and more accurate recognition with reduced computational resources. In this paper, we have explored one of the most promising approaches to HAR utilizing the OpenCV library and Google MediaPipe.By leveraging OpenCV for pre-processing video frames and MediaPipe's 2-step deep neural network detector model to extract landmarks from the human body; we have demonstrated the effectiveness of this approach in achieving desired results. However, despite the progress made, there are still opportunities for further improvement.

Future research efforts should focus on refining the combination of activities and predicting actions that may not exhibit the complete body. This could involve exploring alternative methods for feature extraction and incorporating additional contextual information to enhance the robustness and accuracy of the HAR system.

Overall, the integration of OpenCV and MediaPipe holds great promise for advancing the field of HAR, paving the way for more efficient and reliable human activity recognition systems in various domains such as surveillance, healthcare, and human-computer interaction.

VI. REFERENCES

[1] Mst. Alema Khatun; Mohammad Abu Yousuf; SabbirAhmed; Md. Zia Uddin; Salem A. Alyami; SamerAlAshhab,Deep CNN-LSTM With Self-Attention Modelfor Human Activity Recognition Using Wearable Sensor,https://ieeexplore.ieee.org/document/9781408,IEEE Access paper-2022.

[2] Cuong Pham, Nguyen Thai Son,Huy Tran-Quang,Son Tran,SensCapsNet: Deep Neural Network for Non-obtrusive Sensing Based Human Activity Recognitionty,IEEE Access paper-2020.

[3] Bo Wu; Ting Jiang; Jiacheng Yu; Xue Ding; Sheng Wu; Yi Zhong,Device-Free HAR With Identity-Based Transfer Mechanism, https://ieeexplore.ieee.org/document/9417373,IEEE Access paper-2021.

[4] Quoc V. Le quocle, Marc'Aurelio Ranzato ,Rajat Monga ,Matthieu Devin, Kai Chen ,Greg S. Corrado,Jeff Dean ,Andrew Y. Ng – Building High-level Features Using Large Scale

[5] Mir Kanon Ara Jannat,Md Shafiqul Islam,Sung-HyunYang,Hui Liu,Efficient Wi-Fi-Based HAR Using Adaptive Antenna Elimination, https://www.researchgate.net/publication374244149\,IEEE Access paper-2023.

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

Novel Approach for Detection and Treatment of Anemia and Malaria

Anu Jose[1], Ananya S[2], Aparna C R[2], Megha Sara Paul[2], Varsha Cleetus[2]
[1]Assistant Professor, Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Ernakulam, Kerala, India
[2]Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Ernakulam, Kerala, India

ABSTRACT
This survey explores the pivotal role of Convolutional Neural Network (CNN) frameworks in revolutionizing the timely identification of anemia and malaria, two widespread and critical health challenges. Anemia, characterized by a deficiency in red blood cells, and Malaria, a life-threatening mosquito borne disease, pose significant global health risks, particularly in resource limited regions. The paper begins by outlining the substantial impact of these conditions on public health and underscores the need for efficient diagnostic tools. Subsequently, it comprehensively reviews a spectrum of CNN techniques mainly EfficientNet pretrained model is employed for detection, ranging from traditional statistical methods to advanced deep learning architectures. By synthesizing recent advancements, challenges, and breakthroughs, this survey provides valuable insights into the burgeoning field of CNN driven diagnostic tools, shedding light on their potential to enhance early detection, precision, and accessibility in combating anemia and malaria

Keywords: CNN, EfficientNet, Anemia, Malaria

I. INTRODUCTION

Anemia and malaria stand as formidable adversaries to global health, affecting millions and posing significant threats, particularly in resource-constrained regions. With over 700 million people worldwide grappling with low haemoglobin levels, anemia emerges as a pervasive health concern, with pregnant women and children in developing countries being particularly vulnerable. This condition, marked by a reduction in haemoglobin levels below the WHO-defined normal range, impedes the efficient transport of oxygen throughout the body, leading to various health complications. The current diagnostic landscape involves labour-intensive procedures, including blood sample analysis, complete blood count, and specialized tests, often requiring sophisticated laboratory facilities. Simultaneously, malaria, a life-threatening disease claiming approximately 445,000 lives annually, demands swift and accurate detection. Transmitted through Anopheles mosquitoes, the Plasmodium parasites responsible for malaria wreak havoc by breaking down red blood cells, compromising haemoglobin levels, and impeding oxygen delivery. In response to these challenges, this paper introduces a comprehensive survey on the alternative diagnostic methodology that employs high-resolution blood sample images from

diverse datasets. Leveraging Convolutional Neural Networks (CNN) eliminates the dependency on standard protocols like Complete Blood Count (CBC), paving the way for faster and cost-effective diagnoses. The CNN model, meticulously designed and implemented using Python, undergoes rigorous training on a multitude of microscopic images. Subsequent testing on new samples facilitates the classification of normal blood cells, malaria-infected cells, and distinct types of anemia, namely Megaloblastic Anemia, Sickle Cell Anemia, and Thalassemia.

The objective of this technical exploration is to present an alternative diagnostic methodology that not only circumvents the constraints of conventional laboratory tests but also offers a rapid, cost-effective, and automated solution. The subsequent sections delve into the intricacies of the methodology, CNN configuration, data normalization, and the training and testing phases, providing a comprehensive understanding of the technical nuances underpinning this innovative diagnostic paradigm.

The subsequent sections of the document are organized in the following manner: Section II talks about various system architectures reviewed to create this detection system. Section III discusses a novel approach for detection and treatment of anemia and malaria. Section IV presents the comparative study of various models.

II. LITERATURE SURVEY

Transfer Learning

In [1], a diagnostic system utilizing artificial intelligence to identify malaria parasites within stained blood cell images. The suggested system employs transfer learning for enhancement of the performance of deep learning models in malaria diagnosis as being mentioned in [2] [5]. Specifically, the research employed transfer learning models including VGG19, VGG16, and Inception-ResNet for diagnosing malaria.

The suggested system attained a 95% accuracy in identifying malaria parasites within stained blood cell images. Among the models utilized, the Inception-Resnet model demonstrated the highest accuracy at 95%, followed by VGG16 with 92%, Inception with 93%, and VGG19 with 91%. These favourable results highlight the superior performance of this approach compared to current methods. Additionally, this proposed method holds significance for healthcare professionals involved in screening activities. One of the main merits of the proposed system is its high accuracy in detecting malaria parasites in stained blood cell images. Utilizing transfer learning led to an enhancement in performance of deep learning models in malaria diagnosis, which is particularly useful when there's a shortage of labelled data, particularly in fields like biological image recognition, the suggested system remains broadly applicable and functional across different datasets.

Convolutional Neural Network

This is [13] an innovative approach for the diagnosis and differentiation of Malaria and various types of Anemia, including Megaloblastic Anemia, Sickle Cell Anemia and Thalassemia. The proposed system utilizes Convolutional Neural Networks (CNNs) to process high-resolution blood sample images, eliminating the need for traditional Complete Blood Count (CBC) tests. The CNN, implemented using Python [12], undergoes training on a diverse set of microscopic images sourced from multiple datasets. The trained model is then tested on new images, classifying them into distinct categories: normal blood cells, Malaria, Megaloblastic Anemia, Sickle Cell Anemia or Thalassemia. The proposed method offers a cost-effective and rapid alternative to

conventional laboratory procedures, presenting promising results for expedited diagnosis without the need for specialized analysis laboratories.

The CNN emerges as a key player in this diagnostic system. Comprising 15 layers, including convolutional layers, a flattening layer, max-pooling layers, and fully connected layers, the CNN efficiently extracts and processes features from high-resolution microscopic images. The CNN's architecture incorporates dropout layers to prevent overfitting, padding filters for edge preservation, and activation functions, including rectified linear (ReLU) and Softmax, for maintaining nonlinearity and interpreting output probabilities. The study presents a robust diagnostic system for Malaria and various types of Anemia, leveraging the power of deep learning and CNNs. The proposed system achieves a high accuracy rate and demonstrates potential for further improvement with increased data and higher resolution images. The research contributes to the evolution of diagnostic methodologies, offering a faster and cost-effective alternative to traditional laboratory-based tests.

Double Hidden Layer Extreme Learning Machine Algorithm

Here [7] presents a novel approach for the computerized identification of malaria. The proposed approach incorporates Double hidden layer (DELM), ELM, and Convolutional Neural Networks (CNN) as classifiers. It highlights the severity of malaria as a global health concern, emphasizing the need for a reliable and rapid diagnostic infrastructure. Traditional methods, including microscopy and antigen tests, are criticized for their inaccuracies and time-intensive processes. The paper introduces a machine learning-based solution as mentioned in [10] specifically focusing on the ELM algorithm, combined with CNN for feature extraction in accordance with [9].

The highlight of the proposed CNN DELM model in achieving accurate and efficient malaria diagnosis. The method outperforms existing models, providing promising results for both original and modified datasets. The study recognizes its constraints, including possible differences in image colour, and suggests future research involving the use of CNN feature extraction in multi-layer ELM algorithms. The results of experiments illustrate the efficacy and resilience of the model proposed, highlighting its promise for practical applications in malaria detection.

Machine Learning

Here [6] proposes a system that uses image processing alongside machine learning methodologies to automate the identification and categorization of sickle cells within microscopic images. The system as in [4] consists of two main stages: pre-processing and segmentation. During the initial processing phase, the microscopic image undergoes conversion to a grayscale format, followed by the application of a median filter to eliminate any noise present. The image is then enhanced to remove small objects and sharpen the image. In the segmentation stage, the Otsu thresholding method [11] is employed to distinguish between the three categories of red blood cells (RBCs).: circular, elongated (sickle cell), and other shapes. Next, the Watershed segmentation method is employed to distinguish the overlapping cells. Following this, features related to geometry, statistics, and texture are extracted from the images. Various machine learning classifiers including random forest, logistic regression, naive Bayes, and support vector machine are then utilized to categorize the cells. The proposed system utilizes several machine learning algorithms for the classification of red blood cells (RBCs) in the context of sickle cell anemia diagnosis.

The effectiveness of the suggested system was assessed using precision, F1 score, recall and accuracy measurements. The results showed that the random forest classifier had the highest accuracy for circular RBCs (95%), elongated RBCs (93%), and other shapes (90%). The naive Bayes classifier had the highest precision for circular RBCs (84%), elongated RBCs (88%), and other shapes (90%). The support vector machine classifier had the highest recall for circular RBCs (92%), elongated RBCs (95%), and other shapes (84%). The logistic regression classifier had the highest F1 score for circular RBCs (98%), elongated RBCs (98%), and other shapes (87%). The overall accuracy of the system was 90%.

III. PROPOSED SYSTEM

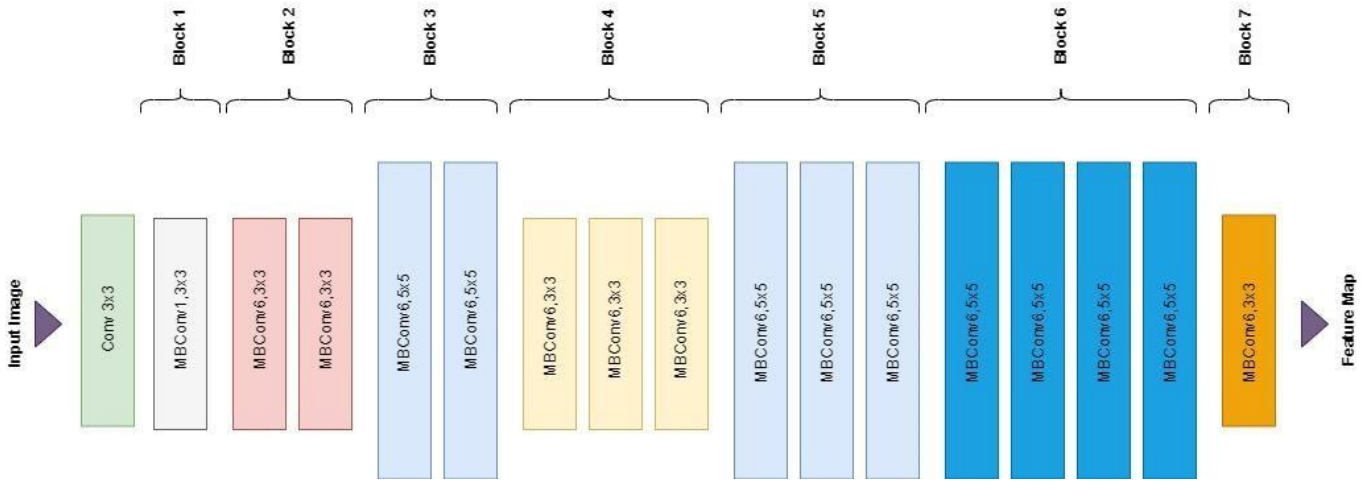


Figure 1: EfficientNet Architecture

Architecture

The proposed system is an advanced Anemia and Malaria Detector leveraging the cutting-edge EfficientNet model for accurate and efficient image classification mentioned in [3]. This innovative system shown in Figure 1 aims to revolutionize medical diagnostics by automating the detection of two prevalent blood-related conditions, anemia and malaria, through the analysis of blood smear and sample images.

At its core, the system employs the EfficientNet architecture, renowned for its superior efficiency in getting high accuracy with a reduced number of parameters. This makes it an ideal choice for resource-conscious applications without compromising diagnostic precision. Leveraging transfer learning, the model is initialized with pre-trained weights on ImageNet, allowing it to capture intricate features crucial for blood-related abnormalities.

The dataset for training and validation is meticulously curated, encompassing a diverse range of annotated images representative of both anemia and malaria cases. Preprocessing steps ensure uniformity and quality, encompassing image resizing and pixel normalization. The EfficientNet model is then fine-tuned on this dataset, adapting its learned features to the specific characteristics of anemia and malaria images. During training, the system incorporates data augmentation techniques to enhance model generalization, and hyperparameters are fine-tuned to optimize performance. The model is evaluated on a separate validation set, ensuring robustness and reliability in real-world scenarios. Rigorous testing validates the system's ability to discern between normal and abnormal blood conditions accurately.

Once validated, the deployed system becomes an indispensable tool for healthcare professionals, offering rapid and precise insights into patients' blood-related disorders. The EfficientNet model's computational efficiency facilitates seamless integration into medical environments, enabling timely and accurate diagnoses. This proposed system stands at the intersection of machine learning and healthcare, promising a transformative impact on the diagnosis and management of anemia and malaria.

A. Implementation

The implementation process began with the collection of labelled datasets comprising blood sample images for anemia detection and blood smear images for malaria detection. After preprocessing the images for consistency, normalization, and resizing, we selected EfficientNet as the model architecture. During training, we monitored the process and applied data augmentation to enhance model generalization. Following successful training, we evaluated the model on a separate validation set, refining it based on performance. Optimization functions are vital for improving the learning process of the model. For instance, the Adam optimizer, which combines elements of both momentum and RMSprop, is commonly used. It helps in adjusting learning rates dynamically, contributing to faster convergence and improved performance. The final trained EfficientNet model was deployed for real-world use, enabling the rapid and accurate identification of anemia and malaria from input images.

TABLE I COMPARATIVE STUDY OF LITERATURE SURVEY

Model	Accuracy	Advantages	Disadvantages
Transfer Learning	95%	Improved accuracy in identifying diseases.	Performance may be affected by the quality of the images used for diagnosis.
Convolutional Neural Network (CNN)	93.4%	Accurate classification	Dependence on high-quality images
Double Hidden Layer Extreme Learning Machine Algorithm	99.5%	High Accuracy	Data Quality and Labelling Issues
Machine Learning	90%	Save time and reduce the risk of human error	Requires a large dataset

IV. CONCLUSION

In conclusion, the surveyed literature underscores the pivotal role of CNN techniques in the detection of anemia and malaria. CNN models, including neural networks, decision trees, and ensemble methods, exhibit promising results in leveraging diverse data sources for accurate diagnosis. These techniques contribute significantly to early detection, improving healthcare outcomes for patients with anemia and malaria. Continued research in CNN applications holds immense potential for advancing diagnostic capabilities and enhancing the effectiveness of healthcare interventions for these prevalent diseases.

V. REFERENCES

[1] G. Zaman Khan, I. Ali Shah, Farhatullah, M. A. Hassan, H. Junaid and F. Sardar," Intelligent Systems for Early Malaria Disease Detection in Patient Cells Using Transfer Learning Approaches," 2023 4th International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, Pakistan, 2023, pp. 1-6, doi: 10.1109/iCoMET57998.2023.10099260

[2] Angel Molina, Jose Rodellar, Laura Bold´ u, Andrea Acevedo, Santiago Alf´ erez, Anna Merino, "Automatic identification of malaria and other red blood cell inclusions using convolutional neural networks", Computers in Biology and Medicine, 2021 doi: https: 10.1016/j.compbiomed.2021.104680.

[3] Y. Guan and Z. Wang," Blood Cell Image Recognition Algorithm based on EfficientNet,"2022 IEEE International Conference on Mechatronics and Automation (ICMA), Guilin, Guangxi, China, 2022, pp. 1640-1645, doi: 10.1109/ICMA54519.2022.9856192.

[4] I. M. I. Alkafrawi and Z. A. Dakhell," Blood Cells Classification Using Deep Learning Technique," 2022 International Conference on Engineering & MIS (ICEMIS), Istanbul, Turkey, 2022, pp. 1-6, doi: 10.1109/ICEMIS56295.2022.9914281.

[5] S. S. Bobde, S. Shenoy, O. Shete, O. Shinde and H. Jhunjhunuwala," Malaria Cell Image Classification using Autoencoder," 2022 2nd Asian Conference on Innovation in Technology (ASIANCON), Ravet, India, 2022, pp. 1-5, doi: 10.1109/ASIANCON55314.2022.9908823

[6] Y B. Sen, A. Ganesh, A. Bhan, S. Dixit and A. Goyal," Machine learning based Diagnosis and Classification Of Sickle Cell Anemia in Human RBC,"2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 2021, pp. 753-758, doi: 10.1109/ICICV50876.2021.9388610.

[7] M. Omaer Faruq Goni et al.," Diagnosis of Malaria Using Double Hidden Layer Extreme Learning Machine Algorithm with CNN Feature Extraction and Parasite Inflator," in IEEE Access, vol. 11, pp. 4117-4130, 2023, doi: 10.1109/ACCESS.2023.3234279.

[8] J. Mitra, K. Vijayran, K. Verma and A. Goel," Blood Cell Classification using Neural Net- work Models," 2023 2nd International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN), Villupuram, India, 2023, pp. 1-5, doi: 10.1109/IC STSN57873.2023.10151543.

[9] K. S. Gill, V. Anand and R. Gupta," An Efficient VGG19 Framework for Malaria Detection in Blood Cell Images,"2023 3rd Asian Conference on Innovation in Technology (ASIANCON), Ravet IN, India, 2023, pp. 1-4, doi: 10.1109/ASIANCON58793.2023.10270637.

[10] G. Prashanthi and S. P. Singh," Identification of Sickle Cell Anemia by Employing Hybrid Optimization and Recurrent Neural Network,"2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN), Salem, India, 2023, pp. 853-858, doi: 10.1109/ICPCSN58827.2023.00146.

[11] I. Verawati and I. D. P. Hasibuan," Artificial Neural Network in Classification of Human Blood Cells Using Faster R-CNN," 2021 4th International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 2021, pp. 86-91, doi: 10.1109/ICOIACT53268.2021.9563974

[12] K. T. Navya, K. Prasad and B. M. K. Singh," Classification of blood cells into white blood cells and red blood cells from blood smear images using machine learning techniques,"2021 2nd Global Conference for

Advancement in Technology (GCAT), Bangalore, India, 2021, pp. 1-4, doi: 10.1109/GCAT52182.2021.9587524

[13] E. W. Abdulhay, A. G. Allow and M. E. Al-Jalouly," Detection of Sickle Cell, Megaloblastic Anemia, Thalassemia and Malaria through Convolutional Neural Network," 2021 Global Congress on Electrical Engineering (GC-ElecEng), Valencia, Spain, 2021, pp. 21-25, doi: 10.1109/GC ElecEng52322.2021.9788131.

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

Multimodal Content Processing across Media Sources

Anila Paul[1], Abhiram Shibu[2], Adarsh Binoy Joseph[2], Anooj Thomson[2], Kevin Sebastian[2]
*[1]Assistant Professor, Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Vazhakulam, Ernakulam, Kerala, India
[2]Department Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Vazhakulam, Ernakulam, Kerala, India

ABSTRACT
An innovative approach is outlined for improving summarization tools, integrating various components seamlessly. By employing advanced NLP techniques and ASR (Automatic Speech Recognition), the system effectively refines text during pre-processing stages. It then utilizes TextRank to generate summaries with proper punctuation and employs a pre-trained BART model for content without punctuations. This method is designed to streamline and enhance summarization processes, specifically tailored to the ever-changing landscape of online multimedia content.
Keywords: ASR, NLP.

I. INTRODUCTION

In the rapidly evolving realm of digital content consumption, there is a growing need for efficient distillation of essential insights from videos. With the vast proliferation of online resources, users often find themselves challenged by the overwhelming amount of information available. This approach aims to simplify content consumption by offering a comprehensive solution for effective video summarization.The central objective of this approach is to enhance the accessibility and utility of video content. Through the seamless integration of visual and audio elements, the system employs advanced natural language processing and machine learning techniques to generate concise text summaries. The motivation behind developing this approach lies in alleviating common issues such as time constraints and information overload faced by individuals engaging with digital content.

Recognizing the significance of videos as valuable sources of information, this adapts its methodology based on the availability of transcripts or the necessity for audio extraction. Its user-friendly design ensures a hassle-free experience for those seeking to distill key information from diverse content sources. This system sets the stage for exploring its potential, emphasizing its ability to revolutionize how users interact with and extract valuable insights from digital content. The system employs an approach utilizing automatic speech recognition, BART[1], natural language toolkit[2], ffmpeg[3], and T5[4] models for the summarization process. With a focus on efficiency, accessibility, and adaptability, this aims to redefine the landscape of digital content consumption.

With a core focus on efficiency, accessibility, and adaptability, this system seeks to redefine the landscape of digital content consumption, offering users a more streamlined and effective approach to engaging with diverse multimedia sources.

II. LITERATURE SURVEY

Introducing an innovative two-step method for video summarization, [1] addresses the challenge of lengthy videos in today's information-heavy environment. Initially, a CNN-based Automatic Speech Recognition (ASR) model accurately transcribes video dialogue into text, ensuring precise textual representations despite accent variations. This phase also integrates a pre-trained model like BART, facilitating transcript generation in cases where pre-existing transcripts are unavailable.

In the subsequent phase, an NLP-based Textrank algorithm, coupled with the pre-trained BART model, condenses the transcribed text into concise summaries by identifying and ranking key phrases. The system's effectiveness is gauged through a comparison of the lengths of the original text and the summary. Moreover, the integration of BART enables the system to dynamically adapt to punctuation presence in the transcripts, enhancing its flexibility to accommodate diverse linguistic styles and dynamic information landscapes. This two-step approach offers an efficient solution for extracting valuable content from videos, enabling users to effectively manage their time amidst the abundance of information.

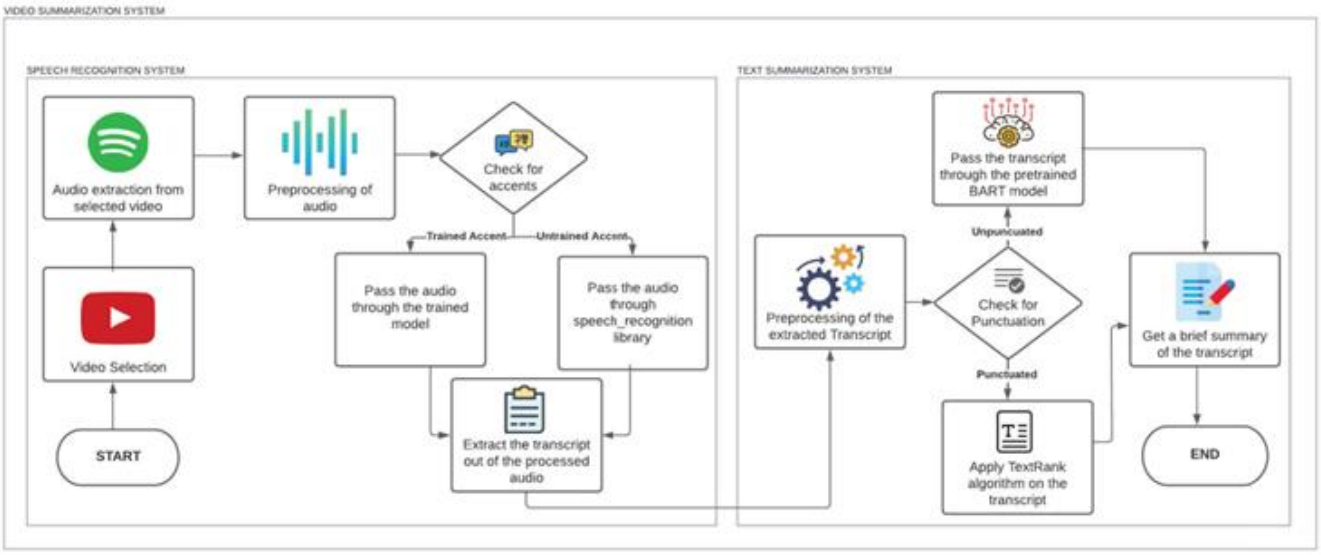


Fig 1. Proposed Methodology

Figure 1: Architecture diagram

In today's digital landscape, technology drives societal advancement, but the sheer volume of online content presents a significant challenge. With platforms like YouTube witnessing a daily upload of approximately 3.7 million videos, users face difficulties in finding relevant information amidst clickbait and overwhelming choices. Summarization skills, focusing on key concept identification and content filtering, are becoming increasingly crucial. The model[2] discussed in this paper offers a solution by transforming summarized text

language, aiding users in quickly assessing a video's relevance. While the paper emphasizes abstractive video summarization, it's worth noting that other approaches, such as extractive summarization, are also explored.

Effective summarization involves several preprocessing steps. Firstly, text cleaning using the Spacy library helps segment text into words and punctuation, facilitating further processing. Sentence tokenization, carried out through tools like the Punkt Sentence Tokenizer, enables the identification of sentence boundaries, crucial for subsequent analysis. Additionally, word tokenization, stemming, and lemmatization refine language comprehension, enhancing the summarization process. Subsequently, summarization techniques prioritize important content by calculating word frequencies and selecting sentences based on the occurrence of key terms. Finally, grammar and spelling checks ensure the quality of the summarized content, contributing to a more effective and reliable summarization outcome

Videos are widely consumed both online and offline, posing a challenge in creating unbiased universal summaries. Summarizing content like a soccer match may vary based on different perspectives, with coaches focusing on techniques while viewers prioritize goals. The overwhelming abundance of videos underscores the need for efficient video summarization techniques.Themodel[3], Utilizing deep neural networks, a model extracts crucial video segments and undergoes evaluation on benchmark datasets such as TVSum, addressing experimental challenges to ensure effective summarization and provide concise outlines for viewers.To achieve effective video summarization, various tools and techniques are employed. FFmpeg facilitates precise segment extraction based on specified start and end times, crucial for focused video summarization tasks. AssemblyAI offers advanced automatic transcription and captioning services, essential for accurately summarizing key elements in videos. By integrating an API key into the workflow, significant time and resources are saved compared to manual review. AssemblyAI's deep learning models continuously improve accuracy, particularly beneficial for scenarios with multiple speakers, enhancing comprehension and contributing to more accurate and informative video summaries. Additionally, the implementation of Natural Language Processing (NLP) is integral to video summarization, extracting key information from audio transcripts to identify significant elements such as keywords, concepts, and topics. NLP corrects grammatical errors and enhances summarization accuracy, highlighting its value in streamlining the process and improving the quality of video summaries.

The headline generation process is revolutionized by combining fine-tuned transformer models, including BART, ProphetNet, and T5, with a dedicated popularity prediction model. This dual-model system [4] seeks to balance accuracy and alignment with ground truth data, enhancing both summarization and headline appeal prediction.Tokenization is crucial in training models, introducing a "headline:" prefix to distinguish headline generation from other tasks.

Fine-tuning involves iterative adaptation with a greedy approach and validation dataset evaluation. The generation phase employs multinomial sampling for diversity, and the final selection is made by a popularity prediction model. This dual-model system effectively captures article essence and predicts audience preferences in a streamlined manner

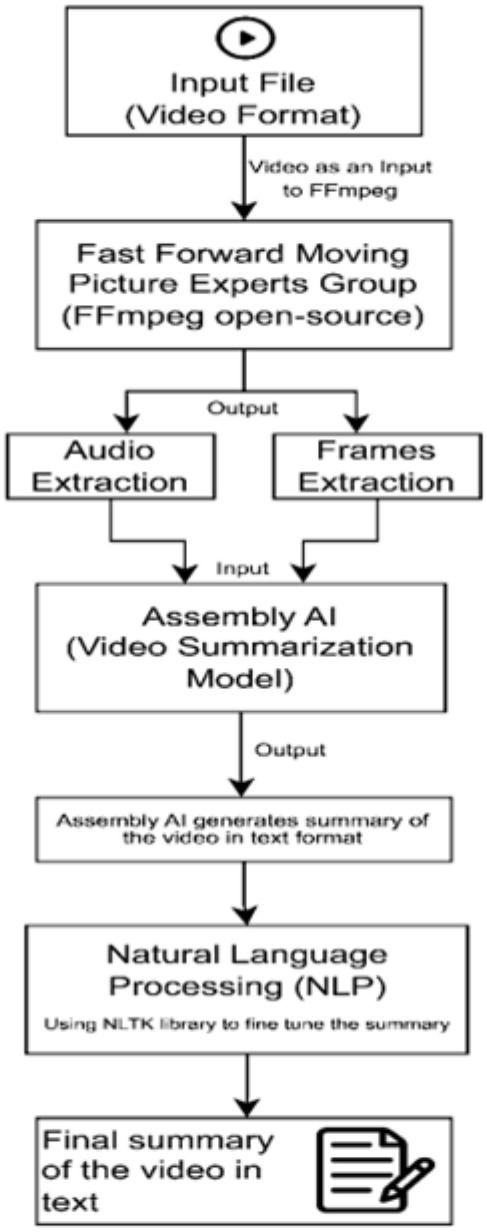


Fig 3. FFMPEG and Assembly AI

Figure 2: Architecture diagram

III. PROPOSED SYSTEM

The proposed system is specifically designed to generate concise textual summaries for given videos. It leverages sophisticated natural language processing and machine learning techniques for efficient execution. In practical usage, the user initiates the system by providing the video they wish to summarize. Upon receiving the input, the system systematically processes the video, checking for the presence of a transcript. If a transcript is available, the system employs Optical Character Recognition (OCR) techniques to retrieve it. Conversely, when

no transcript is accessible, the system resorts to extracting audio from the input video using tools like MoviePy. The extracted audio is then converted into text through Automatic Speech Recognition (ASR) techniques. After the initial processing, a predefined model conducts essential operations, including tokenization, normalization, and sentence segmentation on the obtained text. Subsequently, a secondary processing step evaluates the text for the existence of punctuation. In the presence of punctuation, the system utilizes the TextRank algorithm to prioritize words based on their importance, facilitating an effective summarization process. In cases where punctuation is absent, the system seamlessly employs a pre-trained BART model for the summarization task. This systematic and adaptable approach ensures the system's capability to handle diverse scenarios, providing accurate and coherent textual summaries for a wide range of input videos.

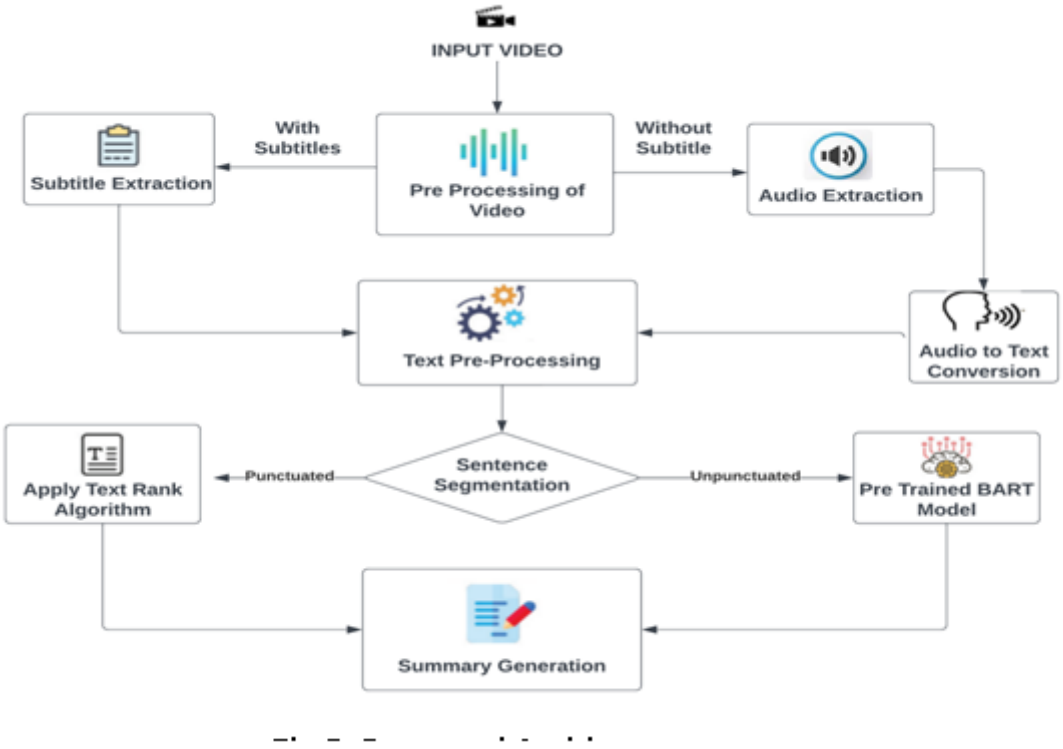


Figure 2: Component Architecture

A. Architecture

The system follows a systematic process for video summarization. It starts by confirming transcript availability using OCR; if transcripts are unavailable, audio is extracted through MoviePy, and ASR converts it to text. Pre-processing ensures text consistency by handling variations, with the model recognizing sentence boundaries. Summarization strategies differ: TextRank prioritizes essential sentences, and BART generates coherent summaries. This method guarantees the production of concise and meaningful summaries from a variety of video sources.

B. Implementation

Initially, the system analyses both visual and auditory components to identify transcript availability. If a transcript is present, Optical Character Recognition (OCR) techniques are applied to extract and convert the text content, making it machine-readable for subsequent summarization. In the absence of transcripts, audio

extraction is performed using tools like MoviePy, separating audio components from video files. Text extraction from audio involves Automatic Speech Recognition (ASR) to convert spoken language into written text. The extracted text undergoes pre-processing using a pre-trained model, addressing variations such as case normalization, stemming, and lemmatization for consistency and clarity. The pre-trained model also identifies sentence boundaries and assesses punctuation presence, crucial for subsequent processing.

Finally, the system employs different approaches for summary generation. For text with punctuation, the TextRank algorithm evaluates sentence significance based on natural language structure. In cases without punctuation, the BART model excels in understanding semantics, crafting coherent summaries without relying on sentence extraction. This comprehensive approach ensures the extraction of key information from videos, providing users with concise and meaningful summaries.

IV. COMPARATIVE STUDY

The below table shows the comparison between the different models.

Paper	*Model*	*Advantage*	*Disadvantage*
Video Summarization using Speech Recognition and Text Summarization	ASR and BART	Enhanced Accessibility, Multilingual Support, Abstractive Summarization.	Accuracy challenges and resource intensive.
Abstractive Summarizer for YouTube Videos	NLTK and spaCy	Versatility and pre-trained.	Resource intensity, slower compared to other models.
AI-based Video Summarization using FFmpeg and NLP	FFmpeg	Cross platform compatibility.	Command - line interface, lack of GUI.
Learning to Generate Popular Headlines	T5	Transfer learning and scalability.	Interpretability and large model size.

V. CONCLUSION

In the ever-evolving realm of digital content, a valuable tool emerges to meet the increasing demand for efficiently extracting insights from a diverse array of video materials. With the continuous growth of online content, users face challenges such as information overload, time constraints, and diverse learning preferences. The system, designed to seamlessly merge visual and audio components, provides a user-centric solution to these challenges. Through the application of advanced natural language processing and machine learning techniques, the system empowers users to distil concise and meaningful text summaries from videos. The adaptability of the methodology, whether working with transcripts or employing audio extraction, underscores the system's versatility in meeting diverse user needs. This research contributes to enhancing the accessibility and utility of video content across various domains. The system not only streamlines the content exploration

experience but also supports the development of personalized and effective engagement. Looking ahead, the system serves as a foundation for further advancements in content exploration technology, emphasizing efficiency, accessibility, and adaptability. As the landscape of digital content continues to transform, the system exemplifies the potential of integrating audio and visual information to redefine how users interact with and extract value from digital content.

VI. REFERENCES

[1] Tirath Tyagi, Lakshaya Dhari, Yash Nigam andRenuka Nagpal, “Video Summarization using Speech Recognition and Text Summarization” IEEE 2023.

[2] Sulochana Devi(B), Rahul Nadar , TejasNichat , Alfredprem Lucas "Abstractive Summarizer for YouTube Videos" 2023.

[3] Hansaraj Wankhede, R Bharathi Kumar, SushantKawade, AshishRamtekkar, Rachana Chawke, "AI-based Video Summarization using FFmpeg and NLP" International Journal of Innovative Science and Research Technology 2023.

[4] Amin Omidvar, Aijun AN, "Learning to Generate Popular Headlines" IEEE 2023.

[5] Ahmed Emad, Fady Bassel, Mark Refaat, Mohamed Abdelhamed,Ashraf AbdelRaouf “Automatic Video summarization with Timestamps using natural language processing text fusion” IEEE 11th Annual CCWC 2021.

[6] MudasirMohd, Nowsheena, MohsinAltafWani, Hilal Ahmad Khanday , Umar Bashir Mir “Semantic-Summarizer: Semantics-based text summarizer for English language text” Elsevier 2023.

[7] Junqing Fan, XiaorongTian, ChengyaoLv, Simin Zhang “Extractive social media text summarization based on MFMMR-BertSum” Elsevier 2023.

[8] Ilampiray, Naveen Raju, Thilagavathy, Mohamed Tharik, MadhanKishore , Nithin A , Infant Raj “Video Transcript Summarizer” E3S Web of Conferences 399, 04015 (2023).

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

NeuraDerm : A Spectrum of Approaches for Skin Disease Detection and Classification

Lakshmi Suresh[1], Ariane Vincent C[2], Georgy P Johnson[2], Jobal Varghese[2], Samuel Kuruvilla[2]
[1]Assistant Professor, Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Ernakulam, Kerala, India
[2]Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Ernakulam, Kerala, India

ABSTRACT
This exploration delves into recent advancements in automating the detection of skin diseases, with a focus onleveraging image processing, Convolutional Neural Networks (CNNs), Support Vector Machines (SVM) and other suitabletechniques for skin disease detection. The urgency for early detection, particularly in the context of potentially lethal conditionslike skin cancer, is underscored. The reviewed methodologiesencompass diverse strategies for classifying skin diseases, including melanoma, incorporating key elements such as imageaugmentation, transfer learning, morphological filtering, anddeep neural networks. The significance of computer vision isscrutinized, comparing the effectiveness of machine learning anddeep learning approaches. Our project aligns with these currents,seeking to contribute to the field by employing image processing,CNNs, SVM, and LBP for the accurate and timely detectionof various skin diseases. This survey serves as a roadmap,offering nuanced insights into methodologies and approaches inthe automated detection landscape.
Keywords: CNN, AlexNet, LIME, VGG, SVM, ResNet, LBP

I. INTRODUCTION

Skin ailments, which range from common afflictions to severe cancers, pose a significant global health challenge. Timely and accurate detection of these conditions is crucial for effective treatment. Recent advancements in automating skin disease detection, using image processing and deep learning, promise to improve diagnostic precision and efficiency. This survey aims to provide a comprehensive overview of the dynamic landscape in automated skin disease detection.

The prevalence of skin conditions, coupled with the severity of issues like skin cancer, highlights the need for robust automated diagnostic systems. Researchers are using Convolutional Neural Networks (CNNs) [1], image processing, and machine learning techniques to classify and diagnose various skin conditions. This survey summarizes key insights from existing literature, including methodologies, challenges, and noteworthy outcomes in the pursuit of automated skin disease detection.

The reviewed literature covers a wide range of techniques, from fusing image augmentation and transfer learning to exploring morphological filtering and machine learning classifiers.We had studied various research papers such asaresearch [2] that mentioned about proposed methodologies covering data augmentation, feature extraction using the ResNet18 algorithm, and prediction explainability using the LIME framework. The experimental analysis section presents the results of the study, demonstrating the effectiveness of the developed model in accurately classifying skin lesions with high accuracy, precision, recall, and F1 score. We have also obtained information from various sources such as a comparative study[3] proposing three different methods for binary classification (melanoma and nevus) using smartphone images, which has been an area lacking in research. The methods rely on different versions of Convolutional Neural Network (CNN) architectures, specifically Alex-net [4], Mobilenet-V2, and Resnet-50 [5]. The models used in the study include Alexnet, Mobilenet-V2, and Resnet-50, each applied with transfer learning [6]and data augmentation. Additionally, the survey examines the role of computer vision in skin disease detection, distinguishing traditional machine learning approaches from advanced deep learning models. By understanding the current state of research in this area, our survey aims to contribute to collective knowledge, guide future research, and be a valuable resource for dermatological health and technology professionals.

II. LITERATURE SURVEY

A. PARALLEL CNN

Convolutional Neural Networks (CNNs) are a class of deep neural networks designed for tasks involving visual data, making them particularly well-suited for image recognition and classification. In the context of skin disease detection, CNNs have proven to be effective in automatically analyzing and diagnosing various skin conditions based on images. Convolutional Neural Networks (CNNs) have become pivotal in the classification of skin cancer lesions. These CNN models have demonstrated remarkable performance, surpassing the capabilities of highly skilled healthcare professionals. This paper [7] proposes a system for the classification of skin cancer.Fig.1 depicts architecture of the proposed system [7].

Fig. 1: System architecture of the proposed model

1) Data Preprocessing and Augmentation: Data preprocessing involves resizing all files to 224 x 224 pixels. Notably, the dataset is split into training, validation, and testing sets while preserving a suitable ratio. It contains photos for various skin diseases.Effective CNN model training requires data augmentation. This technique, performed just-in-time during training, prevents distortion and maintains original data consistency. Choices such as horizontal flip, rotation, shear, and zoom enhance the model's output and address overfitting by diversifying image representations. This approach maximizes the volume of training data, improving the model's ability to accurately classify skin cancer lesions.

The fundamental elements in the system are Feature extraction, Detection and Classification.The architecture contains two phases:

2) Feature Extraction Phase: In Convolutional Neural Networks (CNNs), feature extraction is a crucial process that allows the network to identify and capture important patterns or features in the input data,

such as images. Inorder to extract the features data is processed through several layers are used. Convolution layer, which is a key component in feature extraction. Activation Function is used to introduced non linearity in data and pooling layers to reduce the special dimensions of the data.

3) Classification Phase: The CNN extracts intricate features from input skin images through convolutional and pooling layers, capturing hierarchical patterns indicative of various skin conditions. Following this feature extraction, the classification phase involves flattening the learned representations into a one-dimensional vector and passing them through fully connected layers. These layers make use of the extracted features to classify the input image into specific skin disease categories, facilitating automated diagnosis.

The proposed system produces highest accuracy compared to VGG-16 and VGG-19 models.Also,use of data augmentation techniques increases the volume of data for effective training and improves the model's ability to accurately classify skin cancer lesions.The limitation it mainly faces is the reliance on single dataset,potentially limiting the generalizability of the findings.

B. SVM AND CNN

Skin cancer, including melanoma, is a severe and potentially fatal condition that requires early detection for effective treatment. The traditional methods of diagnosis, such as biopsy and visual examination, can be invasive and timeconsuming. Therefore, there is a growing demand for advanced and efficient techniques to detect skin cancer. The paper [8] includes the framework that focuses on the development and evaluation of machine-learning models for the accurate detection of malignant melanoma skin cancer. Specifically, two classifiers such as Convolutional Neural Networks (CNN) and Support Vector Machines (SVM), are implemented to predict melanoma skin cancer. The goal is to train and test these models using the same data, hardware specifications, and software environment to determine their accuracy and facilitate fair comparisons. The method emphasizes the non-invasive nature of using machine learning models for skin cancer detection and highlights the importance of accuracy and optimization in improving diagnostic capabilities. This method is done using Kaggle's open-source database of 700 malignant and benign skin cancer cases, and the model is used by libraries such as TensorFlow, Keras, pandas and NumPy.

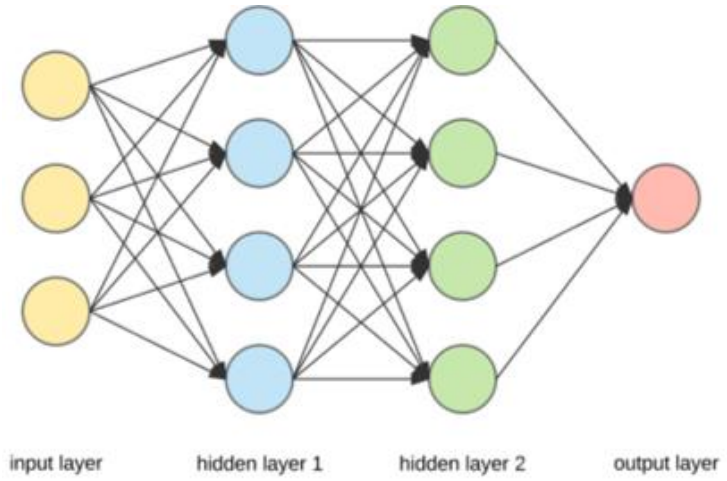


Fig. 2: Convolutional Neural Network

For the CNN model, TensorFlow and Keras libraries can be used. The CNN classifier model is set to sequential, and various layers can be added to the model. These layers include dense, drop, activation, flatten, Conv2D, and

MaxPooling2D. The Conv2D layer has a size of (3, 3), indicating a 3x3 filter for convolution. The MaxPooling2D layer has a pool size of (2, 2), representing a 2x2 pooling operation. The flatten function is added to convert the 2D output into a 1D vector, followed by a dense layer with 64 units. The sigmoid activation function is applied to the output layer. The model is compiled with binary cross-entropy loss and accuracy as the metric of interest. The Fig. 2 shows how different layers are involved in classification in a CNN model.

For the SVM model, the pandas and NumPy libraries are used. The dataset is pre-processed, and individual identity numbers are assigned to the images. The features of the images, such as clump thickness, cell size uniformity, cell shape uniformity, bare nuclei, marginal adhesion, mitosis, normal nucleoli, and single epithelial cell size, are recorded. The data is then separated into training and testing sets. The SVM classifier is fit on the training set and used to predict the classes of the testing set.

Overall, the architecture of the proposed system involved training and testing the CNN and SVM models using the same dataset, implementing specific layers and algorithms for each model, and evaluating their accuracies in detecting malignant melanoma skin cancer.

C. CNN ARCHITECTURE

The paper [9] mainly focus into the methodologies based on CNN for skin cancer classification, emphasizing the effectiveness of CNN models in the diagnosis system. It introduces the HAM10000 dataset, which contains 10015 skin photos intended to train deep learning algorithms, and underlines the issues brought about by the small size and lack of diversity of accessible data. The dataset covers various skin cancer categories and includes images confirmed through histopathology, providing a comprehensive resource for training and testing CNN models. The Fig. 3 represents the basic architecture.

Furthermore, the paper discusses the implementation of CNN models in TensorFlow on the HAM10000 dataset and presents visualizations of the training and testing results, including accuracy, precision, and recall metrics. The model achieved an accuracy of more than 80% with the HAM10000 dataset and tested the model with randomly generated augmented images, obtaining similar accuracy and precision. The paper concludes by highlighting the significance of CNN models in skin cancer diagnosis and references related works in the field.

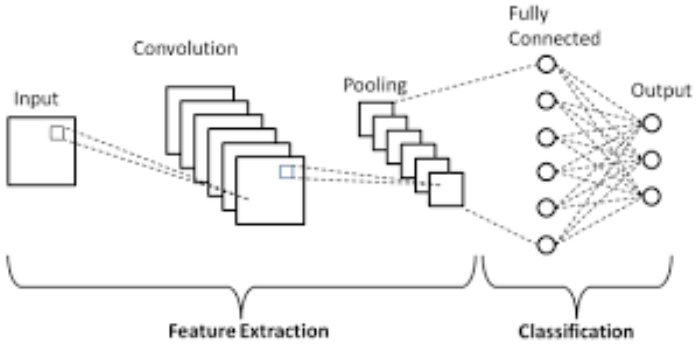


Fig. 3: A simple skin cancer detection using ordinary CNN

In addition, the paper emphasizes the importance of early detection of skin cancer and the potential of computer vision techniques and digital image processing to automate the diagnosis process and reduce manual errors. It also discusses the role of Convolutional Neural Networks (CNNs) in image processing, classification, and object detection, particularly in medical imaging techniques, including lesion classification and tumour diagnosis.

Overall, the paper provides a comprehensive overview of the use of CNN models for skin cancer classification, addressing dataset challenges, data preprocessing, model implementation, and performance evaluation.

D. LBP AND CNN

The paper [10] revolves around building a system using convolutional neural networks (CNN) and local binary models (LBP) and then combining all the points obtained from the CNN and LBP architecture. It uses images as input and checks whether the skin is affected by cancer. The HAM10000 (Human vs. Machine) dataset is used for training and validation of the model. The basic LBP operator is shown in Figure 4.

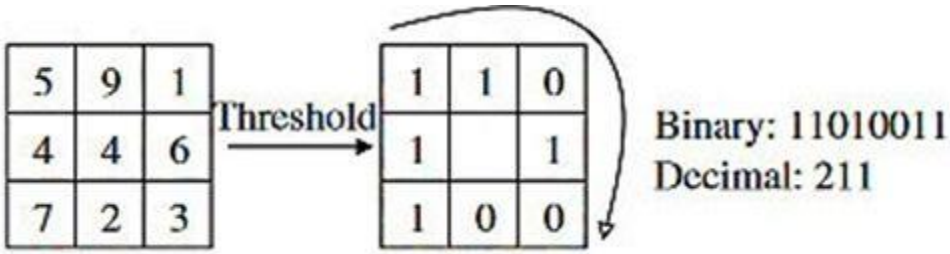


Fig. 4: The Basic LBP Operator

1) Local Binary Pattern: The LBP operator is considered a good tool. The pixels of the input image are summed with the average value from the beginning of each pixel's 3 3 neighbourhood, and the result is treated as a binary number. The histogram of the label is used as the texture descriptor. This histogram contains information about how local micropatterns (such as edges, centers, and flat areas) are distributed throughout the image. In this way the image is divided into regions. Regional histograms are combined to create an overall description of the image.

The advantage of this method is that when the image is divided into areas or regions, some areas have important information for classification. One thing that is difficult to know from a standard deployment perspective is the size of the cluster. Therefore, more complex classifiers are not needed and nearest neighbours are used instead. This method produces unsatisfactory results compared to other methods.

2) Convolutional Neural Network: The CNN architecture is designed to identify seven types of skin problems depending on the number of layers and their layout, activation function, optimizer, number of training sessions and study. The network uses four convolutional layers and two max-pooling layers, a smoothing layer, and then three thick layers with Relu and SoftMax activation functions.
3) Fusion of LBP and CNN: This combination uses the advantages of CNN in reporting good results and LBP in capturing local texture information to improve the skin classification process. The algorithm creates more models by combining features extracted from the two methods. This combination solves the problem of CNN processing limited data, and LBP helps increase power. The disadvantage of this fusion model is that the abstraction levels in LBP and CNN model architectures are different and the fusion of features may require some reduction algorithms to obtain better results.

III. PROPOSED SYSTEM

Introducing our innovative skin app, Skin Nexa which is seamlessly integrated with a powerful CNN model for precise skin disease identification. Beyond diagnosis, this app goes the extra mile by providing personalized skincare recommendations, facilitating doctor appointments, and offering a comprehensive approach to

dermatological well-being. The user-friendly interface ensures easy navigation, while robust privacy measures guarantee data security. With continuous improvements driven by user feedback, this app is designed to evolve alongside emerging technologies, making it a dynamic solution for global users seeking reliable skin health management. Fig.5 represents the Architecture diagram of the proposed system.

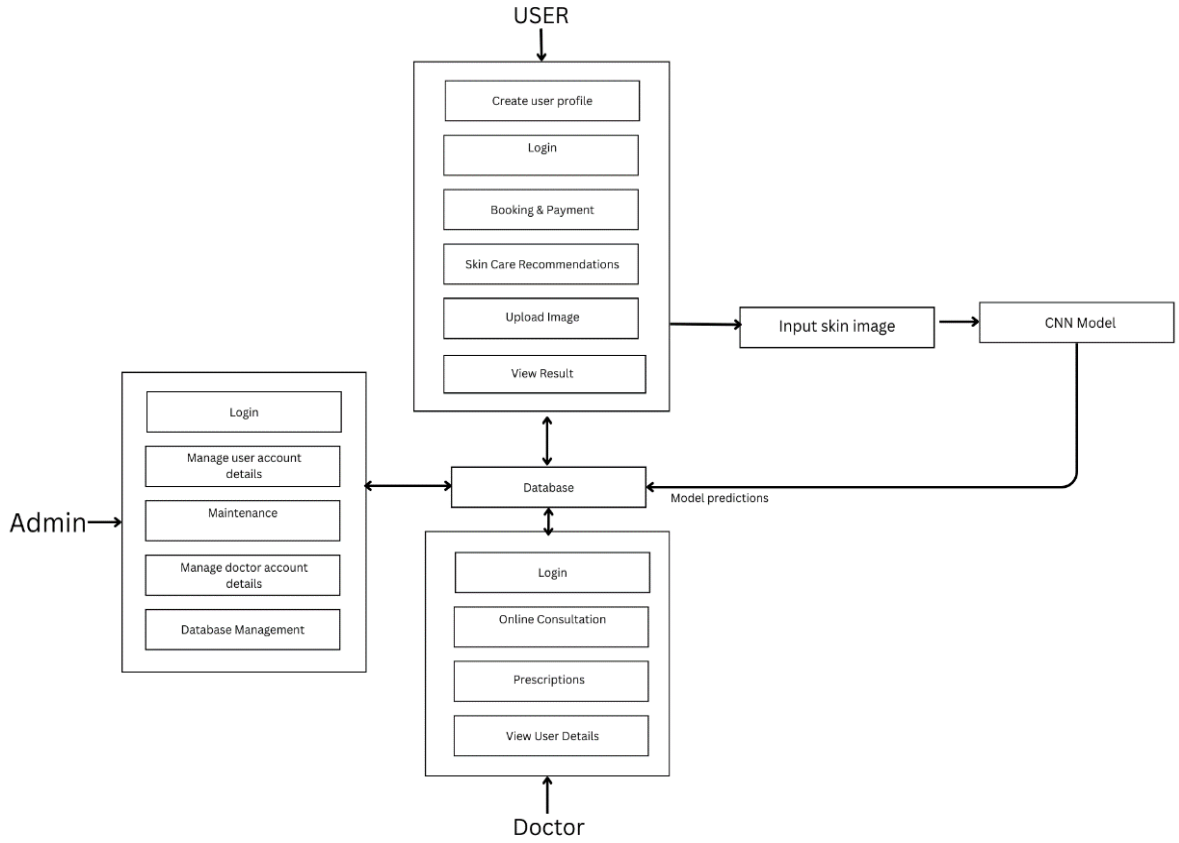


Fig. 5: Proposed System Architecture

Users initiate the process by interacting with the intuitive interface of the 'Skin Nexa' mobile application. Using their smartphone's camera, they capture images of skin abnormalities, initiating the diagnostic process. The user-friendly design ensures accessibility for individuals of varying technical proficiency.

Skin Nexa provides users with instant and accurate diagnoses based on the algorithmic analysis. Beyond diagnoses, the system employs a recommendation engine to offer personalized skincare recommendations. This comprehensive approach equips users with actionable insights, empowering them to proactively manage their dermatological health.

The system incorporates a doctor recommendation component. Based on the identified skin conditions, 'Skin Nexa' recommends specialized dermatologists, fostering a seamless connection between users and healthcare professionals. This feature enhances the user journey, ensuring targeted medical interventions when needed.

IV. COMPARATIVE STUDY

Based on the reviewing of the papers, a thorough comparison analysis was conducted. This study attempted to identify and evaluate the different benefits and drawbacks that are present in every technology that was discussed in the literature. Table 1, which methodically provides a comprehensive summary of the observed

contrasts and captures the complex subtleties of the technologies covered in the corresponding papers, is essential to this analysis. After a methodical examination and evaluation, the table provides significant value as a tool for comprehending the technologies under investigation and provides information about their individual advantages and disadvantages.

Authors	Method	Advantages	Disadvantages
Natasha Nigar, et al	ResNet-18,LIME Framework	Powerful deep neural networks, individual prediction	Complexity, susceptibility to overfitting, and limited interpretability
Sara Medhat, et al.	Alex-ne,MobileNet-V2 ,Resnet50 ,Data Augmentation	high performance , ability to learn complex features	Large model size, highly dependent on data set quality
NoortazRezaona,et al.	Parallel CNN	Detection rate is quite more compared tom other models	Only 9 types of skin cancer can be classified
Christopher Ebuka Ojukwu	SVM, CNN	Accurate Detection Non-invasive Detection, Dataset and Hardware Standardisation	Limited Dataset, Limited Hardware Specification, Lack of Real-Time Implementation
R Raja Subramanian, et al.	Image preprocessing, Image augmentation, Convolution	Use of Convolutional Neural Networks(CNN), High Accuracy, Visualization, Data Preprocessing	Local Optimum,Dataset Limitations, Computational Requirements, Lack of Generalizability Assessment
Lubna Riaz, et al.	CNN ,LBP	Prevents misclassification, CNN + LBP gives high accuracy	not be able to detect other types of skin lesions, merging of CNN & LBP needs dimensionality reduction algorithm

Table 1 :Comparative Study

V. REFERENCES

[1] Tsedenya Debebe Nigat, Tilahun Melak Sitote,Berihun Molla Gedefaw, "Fungal Skin Disease Classification Using the Convolutional Neural Network

[2] Natasha Nigar, Muhammad Umar, Muhammad Kashif Shahzad,Shahid Islam, Douhadji Abalo, "A Deep Learning Approach Based on Explainable Artificial Intelligence for Skin Lesion Classification"

[3] Sara Medhat, Hala Abdel-Galil, Amal Elsayed Aboutabl, Hassan Saleh, "Skin cancer diagnosis using convolutional neural networks for smartphone images: A comparative study"

[4] Mohammad Ali Kadampur, Sulaiman Al Riyaee, "Skin cancer detection: Applying a deep learning-based model driven architecture in the cloud for classifying dermal cell images"

[5] Md.Shahin Ali, Md Sipon Miah, Jahurul Haque, Md.Mahbubur Rahman, Md.KhairulIslam,"An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models."

[6] NoortazRezaoana, Mohammad Shahadat Hossain, Karl Andersson, "Detection and Classification of Skin Cancer by Using a Parallel CNN Model"

[7] Melanoma Skin Cancer Detection Using Support Vector Machines and Convolutional Neural Networks" Christopher Ebuka Ojukwu Department of Computer Science, Nile University of Nigeria, Abuja, Nigeria

[8] R Raja Subramanian, DintakurthiAchuth, P Shiridi Kumar, "Skin cancer classification using Convolutional neural networks"

[9] Lubna Riaz, Hafiz Muhammad Qadir, Ghulam Ali, Mubashir Ali, Muhammad Ahsan Raza, Anca D. Jurcut, Jehad Ali, "A Comprehensive Joint Learning Systemto Detect Skin Cancer"

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

SARS: Mental Health Chatbot Using Natural Language Processing

Remya Paul[1], Abi Mathew Kurian[2], Melvin Johnson[2], Nirmal Vijayan[2], Sebastian Skaria[2]
[1]Assistant Professor, Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Muvattupuzha, Kerala, India
[2]Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Muvattupuzha, Kerala, India

ABSTRACT
In today's digital landscape, where the constant flow of text-based communication shapes our interactions and experiences, "SARS: Mental Health Companion" emerges as a beacon of understanding in the realm of emotions. This project sets out to decipher and comprehend the sentiments expressed in text data, with the ultimate goal of offering meaningful insights and support to individuals navigating the intricacies of their emotional states. SARS utilizes a combination of natural language processing and machine learning techniques to categorize text into distinct emotional categories, ranging from joy and contentment to sadness and anger. The system's innovation lies not only in its ability to detect emotions but also in its capacity to provide personalized recommendations, be it motivational quotes to uplift the disheartened, soothing techniques to alleviate stress, or thought-provoking content for those seeking enlightenment. SARS represents the fusion of technology and human emotions, extending a helping hand to individuals as they navigate the complexities of their inner experiences in our increasingly interconnected digital age.

Keywords— NLP, ML Techniques, Emotions Classification, Digital Landscape, Support

I. INTRODUCTION

In an era facing rising mental health concerns and a growing demand for accessible support, innovative technological solutions are emerging as a promising avenue. One such advancement is the proposed mental health chatbot application, which has the potential to revolutionize the way we address mental health needs within society.

This application envisions a future where individuals struggling with diverse mental health challenges have access to a safe and nurturing environment. It addresses the critical need for accessible and empathetic support, a need underscored by the global mental health crisis. From stress and anxiety to depression and more complex conditions, individuals across demographics face a spectrum of mental health concerns. Unfortunately, accessing timely assistance can be hampered by limited resources, societal stigma, and apprehensions about seeking professional help. This gap highlights the urgent need for a platform that is empathetic, non-judgmental, and readily accessible to cater to varying mental health needs.

The envisioned application steps into this void by offering a discreet and user-centric interface. Its core functionality lies in an empathetic conversational interface driven by advanced Natural Language Processing (NLP) algorithms. By leveraging technology to understand and respond to users' emotional states with compassion, this application aims to become a reliable companion, providing solace and guidance during their mental health journey.

II. LITERATURE SURVEY

A. DECISION TREE AND CART ALGORITHM

Beyond their well-established roles in customer service and e-commerce, chatbots are rapidly transforming the healthcare landscape. Their ability to handle routine tasks like answering FAQs, scheduling appointments, and managing basic health inquiries frees up human healthcare professionals to tackle complex cases and offer personalized care. In e-commerce, chatbots excel at tailoring product recommendations, providing real-time order updates, and resolving customer support issues quickly and efficiently. By providing 24/7 access to health information and guidance, chatbots empower patients to become active participants in their healthcare journey. The paper proposed in [1] utilizes a personalized support that not only fosters informed decision-making but also alleviates pressure on human healthcare providers.

The healthcare chatbot system is structured with a modular design, commencing with the User Interface module serving as the initial interaction point for users. Subsequently, user input undergoes processing by the Natural Language Processing (NLP) module, where meaning and context are extracted before forwarding the formatted data to the Decision Tree module. Using the algorithm, the module analyzes the data to generate suitable responses and may also engage with the Knowledge Base module for precise healthcare information. The decision-making aspect of the Decision Tree considers the structure of the tree, configured hyperparameters, and the knowledge base to discern the most pertinent response for each user query. Additionally, the Personalization module refines responses by integrating outputs from the Decision Tree module and user-specific data, tailoring information for individual users. The architectural framework emphasizes the smooth interconnection among the NLP, Decision Tree, Knowledge Base, and Personalization modules, empowering the chatbot to deliver precise and personalized healthcare responses. Ongoing updates to the knowledge base are imperative for upholding the chatbot's dependability. The modules are illustrated in figure 1.

The Natural Language Processing component is fundamental to the server-side architecture, empowering the system to understand user input by extracting meaning and context. The decision-making module analyzes user inquiries and generates fitting responses, ensuring the Chatbot's effectiveness in providing contextually appropriate information. The knowledge base serves as a rich repository of healthcare-related information, acting asa vital resource for the Chatbot to enhance its understanding and deliver accurate responses. Additionally, the personalization component focuses on tailoring the Chatbot's responses to individual users, incorporating user-specific data and preferences.

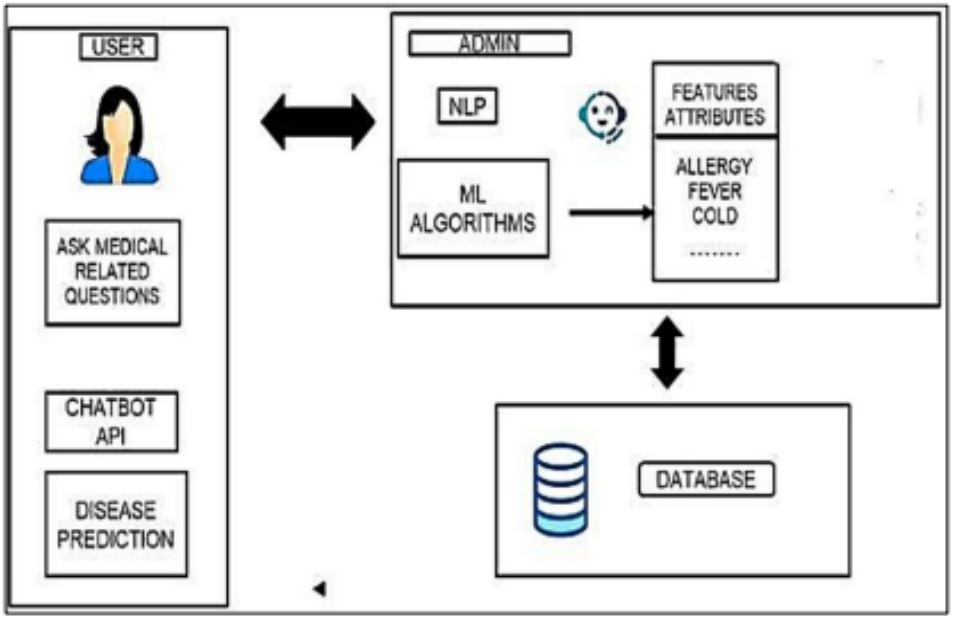


Figure 1: Modules of the helpi chatbot

B. DEEP LEARNING NLP

Amidst the COVID-19 pandemic, extended lockdowns became essential to contain the virus until vaccines could be developed. However, isolation led to heightened levels of anxiety, depression, and other mental health issues. Recognizing this need for support, various studies explored the potential of chatbots [5] for emotion recognition. Despite advancements, limitations persist. The paper proposed in [2] addresses this gap by proposing a chatbot tailored to alleviate pandemic-induced psychological distress, particularly among students. By employing natural language processing and deep learning models, this innovative chatbot aims to understand and address the root causes of mental distress, offering support to regulate emotions and counter negative thought patterns.

The Natural Language Processing (NLP) unit of the system utilizes a deep learning-powered, multi-stage preprocessing pipeline to refine user requests for sentiment analysis effectively. The holistic approach ensures clear, concise data for accurate sentiment analysis and deep learning model comprehension. The Natural Language Comprehension unit, the core of the system, transforms user statements into structured patterns. The method utilizes an LSTM-CRF-based Seq2Seq model to decipher user intents and patterns within textual inputs, empowering the unit to understand user intentions effectively.

The Natural Language Generation (NLG) unit translates abstract meanings into coherent textual responses for users, leveraging machine learning for personalized responses. Unlike rule-based systems, modern NLG incorporates diverse inputs for tailored and relevant outputs, enhancing the conversational experience. The Dialogue Management (DM) unit orchestrates interactions and responses by employing Dialogue State Tracking (DST) and Policy Learning. DST infers the current conversational state, while Policy Learning determines the optimal response, leading to an efficient and goal-driven user experience. This interplay allows the system to track conversation progress, understand context, and deliver relevant responses effectively. The architecture of the model is shown in figure 2.

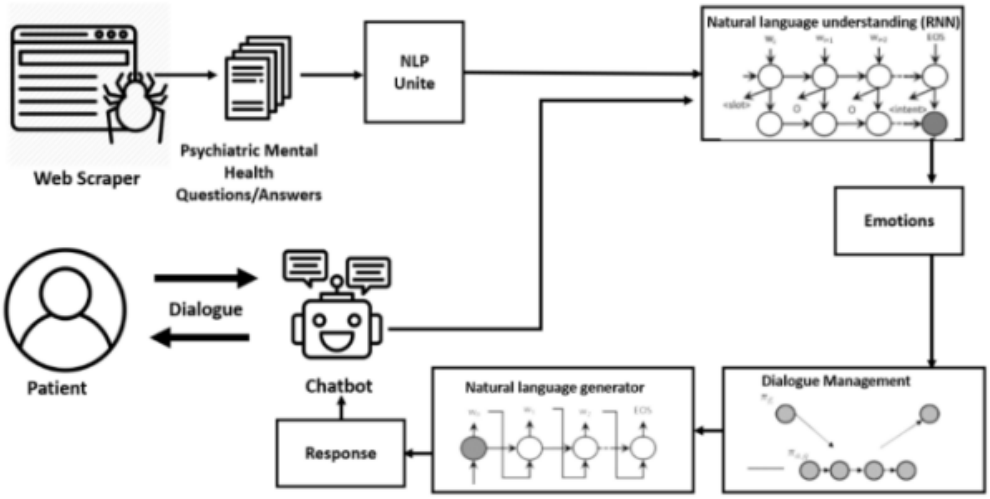


Figure 2: Deep learning based architecture

C. SEQ2SEQ MODEL

As the pandemic's psychological impact continues to evolve, the persistent stigma surrounding mental health discussions fuels stress and hinders productivity [7]. In response to this challenge, a Therapy Chatbot emerges as a beacon of hope. Offering a confidential space free from judgment, this innovative tool acts as a virtual companion, engaging users in empathetic conversations and providing supportive guidance reminiscent of a real therapist. By fostering open expression, offering timely advice, and facilitating access to resources, the chatbot aims to reduce deaths from untreated depression. Its non-threatening, readily available nature encourages proactive mental health care, holding the potential to mitigate the severe consequences of unaddressed mental health issues and contribute to a healthier emotional landscape for all.

The Therapy chatbot as proposed in the paper [3] leverages the power of the Sequence to Sequence (Seq2Seq) model, deeply integrated within the DialogFlow framework. The model acts as the central nervous system, orchestrating personalized conversational sequences addressing individual emotional needs. Whether navigating the depths of grief or navigating career challenges, DialogFlow's intent system as shown in Figure 3, meticulously trained with a tailored dataset, allows the chatbot to effectively identify keywords within user input. This enables the Seq2Seq model to generate relevant and empathetic responses, fostering a dynamic conversational interface that adapts to the unique emotional landscape of each user. Through this intricate interplay of NLP techniques, the Therapy chatbot transcends mere conversation, offering personalized support and fostering a safe space for individuals to explore their emotions [8] and navigate through life's challenges.

The Seq2Seq model forms the cornerstone of the chatbot's preprocessing pipeline, playing a crucial role in user interaction and engagement. The model efficiently breaks down user inputs into meaningful tokens, removing irrelevant information and highlighting key elements. This processed data is stored within the chatbot's dataset, readily accessible for retrieval and presentation during conversations. Utilizing Seq2Seq, the chatbot retrieves relevant responses, ensuring timely and accurate answers tailored to individual concerns. Moreover, the model powers questioning techniques, aiding in identifying user needs and suggesting coping mechanisms. When necessary, the chatbot [6] seamlessly directs users to professional resources, facilitating therapist recommendations and appointment scheduling. This intricate integration between Seq2Seq and DialogFlow

exemplifies the chatbot's empathetic approach, providing personalized support and guiding users through a streamlined conversational journey, ultimately enhancing the overall user experience.

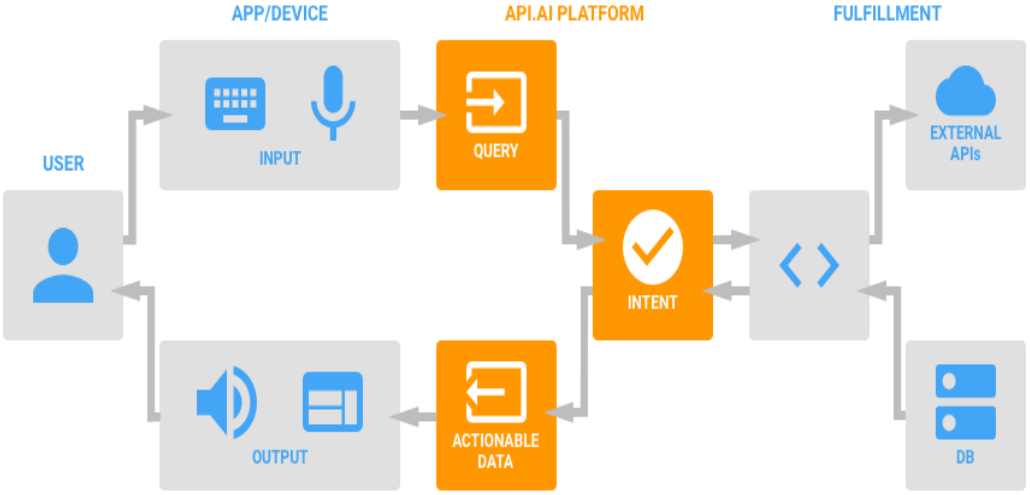


Figure 3: Architecture of DialogFlow

D. BA ORIENTED NLP

The global prevalence of mental illnesses, affecting nearly a billion people, demands urgent attention. Conditions like anxiety and depression escalate, with an estimated annual economic burden of USD 2.5 trillion, projected to reach USD 6 trillion by 2030. The World Health Organization's Special Initiative for Mental Health emphasizes the need to expand access to quality care. Cognitive Behavioral Therapy (CBT) stands out as an effective non-pharmacologic approach. Leveraging innovations like chatbots such as Woebot extends mental health support widely. The paper proposed in [4] suggest that Behavioral Activation (BA) and Artificial Intelligence (AI) in chatbots offer personalized assistance. The paper outlines a research agenda for AI-driven chatbots, marking progress in mental health interventions.

The conceptual framework for the BA-based AI chatbot evolved through a structured three-phase approach. A review of BA adoption and insights from mental health support groups informed the framework. Validation by experts highlighted the chatbot's role as a supportive companion, not a replacement for healthcare services [6]. The framework incorporates continuous personalized interactions and ethical considerations. The framework of BA is shown in figure 4.

Participatory evaluation confirms Bunji's effectiveness in providing mental health support. As conversational agents become prevalent, Bunji shows promise in scaling up to support frontline workers and communities. Future work includes long-term evaluation, creating user communities, and expanding gamification features. Bunji represents a technological leap in chatbots, offering personalized behavioral activation and remote health monitoring, contributing to society's well-being.

Furthermore, the integration of user feedback mechanisms within Bunji's interface allows for continuous improvement and adaptation to evolving user needs. By actively soliciting input from users and mental health professionals alike, Bunji can refine its algorithms and expand its repertoire of supportive interventions. This iterative process ensures that Bunji remains relevant and effective in addressing the diverse needs of individuals dealing with mental health challenges [8]. Additionally, ongoing collaboration with researchers and stakeholders facilitates the incorporation of cutting-edge advancements in AI and behavioral science into Bunji's functionality, positioning it at the forefront of innovative mental health interventions. As Bunji

continues to evolve, it holds the potential to revolutionize the landscape of mental health care delivery, offering accessible and personalized support to individuals worldwide.

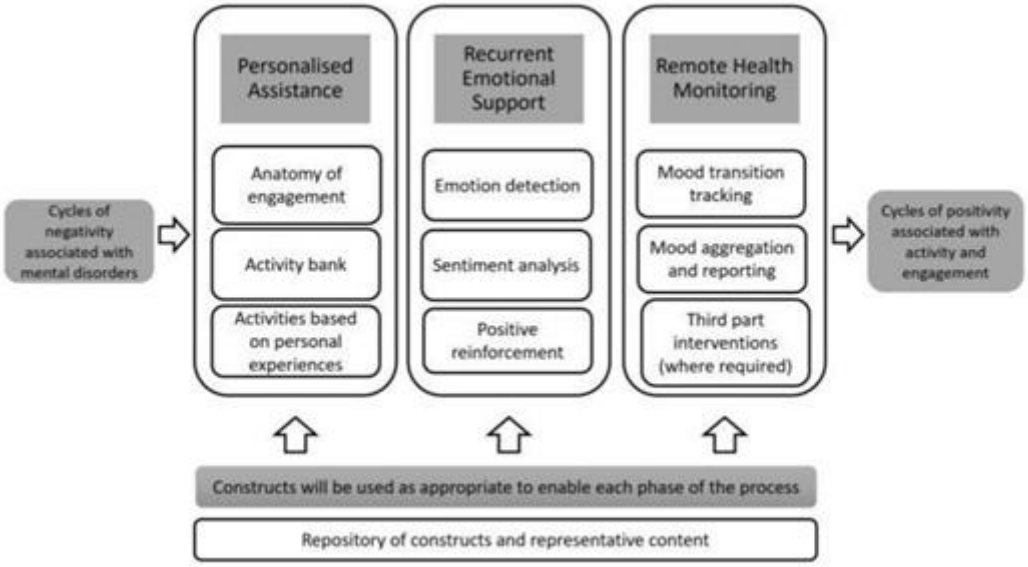


Figure 4: Framework of behavioral activation

III. PROPOSED SYSTEM

A. Architecture

The proposed architecture for the sentiment analysis and recommendation system chatbot integrates various modules to facilitate user interaction, and therapist engagement effectively. Here's a breakdown of the architecture.

Upon initial login, the chatbot collects user preferences such as preferred conversation topics, therapist preferences, and notification settings. This information is crucial for personalizing the user experience and tailoring recommendations.

TALK: This module enables users to converse with the chatbot. Using Natural Language Understanding (NLU), the chatbot comprehends conversation context. Subsequently, Natural Language Processing (NLP) identifies keywords and emotion within user input. The chatbot then generates responses using Natural Language Generation (NLG) techniques. The Dialogue Management (DM) component orchestrates conversation flow, ensuring coherent responses.

JOURNAL: This module tracks user activity logs, including conversation history and sentiment trends. Understanding user behavior and sentiment patterns enables personalized recommendations and interventions.

ADD THERAPIST: Users can schedule sessions with therapists or add preferred therapists to their profile using this module. The chatbot facilitates scheduling appointments and managing therapist preferences, improving user access to mental health support.

The chatbot sends push notifications to users, providing updates, reminders for scheduled sessions, or notifications about new features and content. These notifications enhance user engagement and keep users informed about relevant activities. The architecture in Fig.5 integrates user preferences, sentiment analysis, therapist engagement, and push notifications to create a comprehensive chatbot experience for mental health

support and recommendation. Through NLU, NLP, NLG, DM, and user activity tracking, the chatbot aims to deliver personalized interactions while promoting user engagement and well-being.

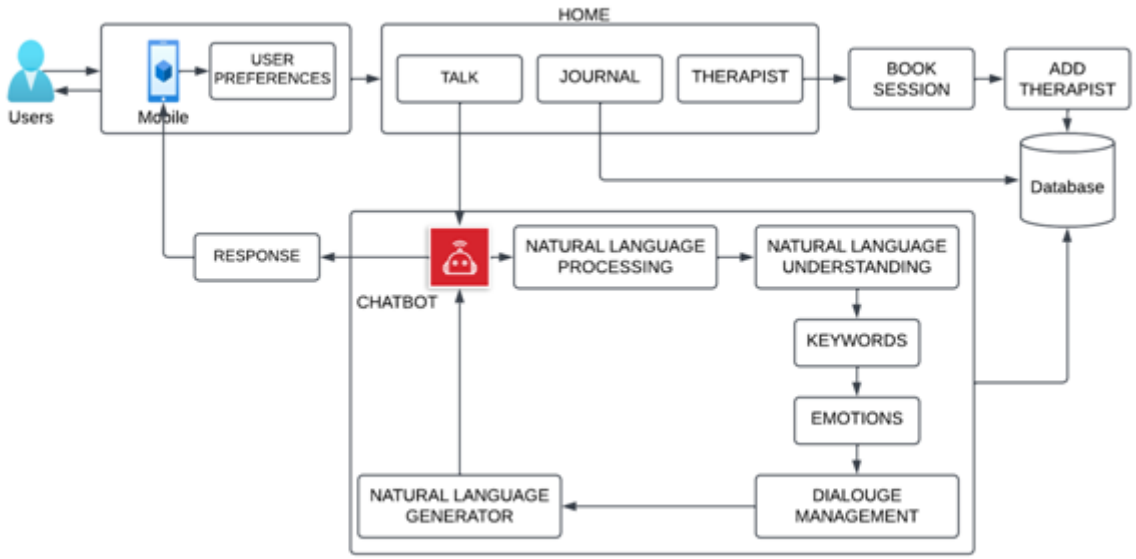


Figure 5: System architecture of SARS chatbot

B. Implementation

Implementing a sentiment analysis and recommendation system chatbot involves several crucial steps to ensure functionality, effectiveness, and user engagement. The process starts with thorough planning and requirement gathering, understanding user needs, preferences, and desired functionalities. This stage includes conducting comprehensive user research, defining clear use cases, and outlining the necessary features and modules for the chatbot to deliver meaningful interactions and support effectively.

Once the planning phase is complete, the technical architecture of the chatbot must be meticulously designed and implemented. This involves setting up the necessary infrastructure for hosting the chatbot and developing the backend logic for user authentication, session management, and data storage. Additionally, integration with natural language processing (NLP) and natural language understanding (NLU) frameworks is crucial. These integrations empower the chatbot to understand user queries, perform sentiment analysis and generate contextually relevant responses, fostering meaningful and empathetic conversations with users.

Subsequently, the main modules of the chatbot, including "TALK," "JOURNAL," and "ADD THERAPIST," need to be developed and seamlessly integrated. Each module requires specific functionality and interaction flows to cater to users' diverse needs. For instance, the "TALK" module necessitates the implementation of conversational logic, sentiment analysis capabilities, and response generation using NLP, NLG, and dialogue management techniques. Meanwhile, the "JOURNAL" module entails logging user activity, sentiment trends, and conversation history to provide personalized recommendations and insights. Lastly, the "ADD THERAPIST" module requires functionality for scheduling therapy sessions and managing therapist preferences, enhancing users' access to mental health support resources and services.

In summary, implementing the sentiment analysis and recommendation system chatbot requires a meticulous and coordinated effort across various stages of planning, design, development, and testing. Prioritizing user-centric design principles and leveraging advanced technologies such as NLP and NLU, the chatbot aims to

provide empathetic and personalized support to users seeking mental health assistance. Through continuous iteration and improvement based on user feedback, the chatbot endeavors to evolve into a valuable tool for promoting well-being and facilitating access to mental health resources and services.

IV. COMPARATIVE STUDY

A comprehensive comparative analysis was undertaken based on the examination of the papers. The study aimed to delineate and assess the various advantages and disadvantages inherent in each of the technologies explored within the literature. Central to this examination is Table 1, which meticulously presents a detailed overview of the observed comparisons, encapsulating the nuanced intricacies of the technologies discussed in the respective papers. Through a systematic review and analysis, the table serves as a valuable resource for understanding the landscape of technologies within the scope of the study, offering insights into their respective strengths and limitations.

Table 1: Comparative study

Author	Technology	Advantage	Disadvantage
G. Karuna, et al.	NLP Processing: Natural Language Processing.	Users can seek information or assistance at any time.	The study sample is small for generalization.
Intissar Sahli, et al.	Natural Language Understanding and Machine Learning Techniques.	Tailored Responses.	Heavily depends on external sources like Reddit for information, potentially affecting the accuracy of its responses.
Pranav Kapoor, et al.	Seq2seq Model Using Dialog Flow(RNN)	Tailored responses and advice.	Limited ability to understand complex emotions beyond predefined keywords.
PrabodRathyanaka, et al.	Natural Language Learning Techniques	Users can provide more information for personalized conversation	The current model may not be equally suitable for all user groups.

V. CONCLUSION

This survey comprehensively explored the landscape of various mental health chatbots and their supporting frameworks, showcasing diverse functionalities and critically evaluating their effectiveness and limitations. The findings highlight the burgeoning potential of chatbots in democratizing mental health support, particularly for individuals facing affordability constraints, geographical barriers, or stigma. By leveraging advancements in artificial intelligence and user-centered design, these chatbots offer a promising avenue for enhancing accessibility, early intervention, and personalized mental health care, paving the way for a more inclusive and

integrated approach to well-being. This, in turn, aligns with the global vision of promoting mental health equity and fostering a more supportive and resilient society.

VI. REFERENCES

[1] G. Karuna, Gomaram Gowthami Reddy, J. Sushmitha, Bitla Gayathri, Sameer Dev Sharma, Debnarayan Khatua, Et al. Helpi - An Automated Healthcare Chatbot [2023]. 10.1051/e3sconf/202343001040.

[2] Intissar Sahli, Kamal El Guemmat, Mohammed Qbodou, Khalifa Mansouri, ET al. Towards developing a pocket therapist: an intelligent adaptive psychological support chatbot against mental health disorders in a pandemic situation[2021]. http://doi.org/10.11591/ijeecs.v23.i2.pp1200-1211.

[3] Pranav Kapoor, Pratham Agrawal, Zeeshan Ahmad, Et al. Therapy Chatbot: A Relief From Mental Stress And Problems [2021]. 10.14299/ijser.2021.05.08.

[4] Prabod Rathyanaka, Nishan Mills, Donna Burnett, Daswin De Silva, Daminda Alahakoon, Richard Gray, Et al. A Mental Health Chatbot with Cognitive Skills for Personalised Behavioural Activation and Remote Health Monitoring [2022]. 10.3390/s22103653..

[5] Falguni Patel, Riya Thakore, Ishita Nandwani, Santosh Kumar Bharti, Et al. Combating Depression in students using an intelligent chatbot: A Cognitive behavioral therapy[2019]. 10.1109/INDICON47234.2019.9030346.

[6] Mounika Karna, Sujitha Juliet D, R. Catherine Joy, Et al. Deep learning based Text Emotion Recognition for Chatbot applications [2020]. 10.1109/ICOEI48184.2020.9142879.

[7] Marion Jojoa, Parvin Eftekhar, Behdin Nowrouzi-Kia, Begonya Garcia-Zapirain. Natural language processing analysis applied to COVID-19 open-text opinions using a distilBERT model for sentiment categorization [2022]. https://doi.org/10.1007/s00146-022-01594-w.

[8] Santosh Kumar Bharth, S Varadhaganapathy, Rajeev Kumar Gupta, Prashant Kumar Shukla, Mohamed Bouye, Simon Karanja Hingaa, Amena Mahmoud, ET al. Text Based Emotion Recognition using Deep Learning Approach [2023]. https://doi.org/10.1155/2022/2645381.

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

Bridging Languages: Unified Speech-to-Speech Translation

Bency Cleetus[1], Arjith Gurudas[2], Edwin Roy[2], Harshavardhan Dhinu[2], Lino Saji[2]
[1]Assistant Professor, Department of Computer Science, Viswajyothi College of Engineering and Technology, Vazhakulam, Kerala, India
[2]Department of Computer Science, Viswajyothi College of Engineering and Technology, Vazhakulam, Kerala, India

ABSTRACT
As multitudes of cultures intermingle with one another over the expansive web of the Internet, the need for accurate and efficient language translations become more and more evident. Traditional methods of speech conversion which involve building a cascaded model of ASR-MT-TTS are inefficient in the terms of the computational power they use and error propagation probability they inherently possess. It is evident that this need for translation will only grow more as technology advances further, making the aspects of accuracy, efficiency and computational power become more important. The reviewed methods contain many such strategies that facilitate such direct speech translation. Alternate related models are compared in terms of aforementioned metrics. Our project also aligns with these as we put forward a direct speech-to-speech translation model that skips the cascaded nature of traditional models and instead on translation without intermediary text representation, along with additional data augmentation.
Keywords: S2UT, S2ST, HuBERT, HiFiGAN, Wav2vec

I. INTRODUCTION

In today's interconnected world, successful cross-linguistic communication is a critical component of cooperation, comprehension, and advancement. The necessity of improving language translation systems is highlighted by the world's increasing demand for seamless translation services in a diverse and culturally interconnected environment.Language limitations continue to be major roadblocks to cross-cultural collaboration and the exchange of ideas as the world grows increasingly interconnected. Conventional translation methodologies frequently entail intermediary stages, such as translating text, which may introduce complications and mistakes. By immediately translating spoken words from one language to another in real time, the suggested direct conversion model aims to transform this procedure. Other methods used for this include transcoding by multitask learning[1]. Auxiliary tasks such as speech separation via conformer[2] were also explored to ensure quality of speech detection for better conversion. In this way, ASR was also improved by using word embedding predictions[3]. Data mixing at three levels[4] also provides a more efficient training that gives state of the art results.As mentioned before, the globalized world needs translations that are both quick and efficient. With an unwavering focus on accuracy, efficiency and effectiveness, this system aspires to

reshape the way we break down language barriers and enable seamless communication across diverse communities.

II. LITERATURE SURVEY

A. DISCRETE UNITS

This paper introduces a direct speech-to-speech translation (S2ST) model that avoids intermediate text generation, presenting a dual-phase approach[5]. The initial phase involves applying a self-supervised discrete speech encoder on the target speech, followed by training a sequence-to-sequence speech-to-unit translation (S2UT) model to predict discrete representations. A joint speech and text training framework is designed for cases with available text transcripts, allowing the model to generate dual modality output in the same inference pass. Experimental results on the Fisher Spanish-English dataset demonstrate a 6.7 BLEU improvement compared to a baseline S2ST model predicting spectrogram features.

The proposed system is a transformer-based model with a speech encoder and discrete unit decoder, incorporating auxiliary tasks for model learning. The speech-to-unit translation (S2UT) model utilizes HuBERT for self-supervised speech representation learning, applying k-means clustering to generate discrete labels. Two strategies for predicting discrete unit sequences, "stacked" and "reduced," are explored. Multitask learning incorporates auxiliary tasks, and a unit-based vocoder, utilizing HiFi-GAN with modifications, converts discrete units into waveforms. The proposed system demonstrates comparable performance to models trained with text supervision, showcasing its potential for translation between unwritten languages. The basic framework is illustrated as Fig 1.

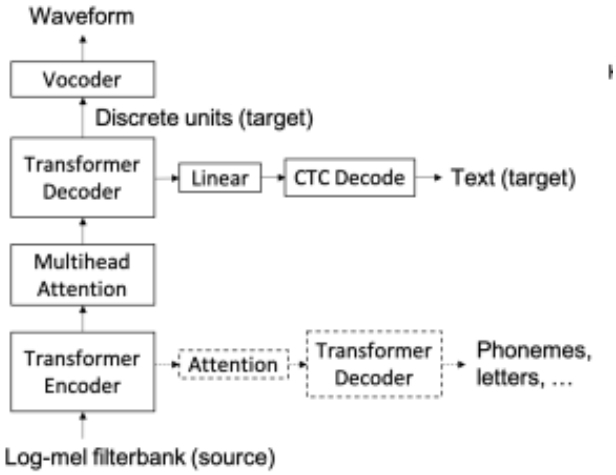


Fig 1. S2UT-Vocoder Framework

B. HUBERT

The paper proposes approach HuBERT which addresses the issues of self-supervised learning of speech levels, by using an offline clustering step to provide aligned target labels for BERT-style prediction loss[6]. HuBERT's unique feature is that it applies missing predictions only to covered regions, forcing the machine to model joint sounds and speech on continuous outputs HuBERT distinguishes itself by its focus a it focuses on the stability of unsupervised clustering steps rather than on the quality of assigned cluster labels. HuBERT uses an acoustic

unit discovery model to assign frame-level targets, while representation learning is performed by a mask prediction model. The masking method is consistent with SpanBERT and wav2vec 2.0 concepts. The analysis uses weighted loss functions that balance the masked and unmasked time steps. The weighting parameter α controls whether covered or uncovered regions are emphasized. The proposed method determines the robustness of group goals as shown in experiments. Furthermore, the paper also explores how cluster models are used to optimize objectives by combining different clustering models to provide useful information for representative learning. A restructuring of cluster guidelines is proposed positions have improved through continuous learning has also occurred.The HuBERT implementation is based on the wav2vec 2.0 framework and comes in three configurations: BASE, LARGE, and X-LARGE, which vary in model size and complexity A convolutional waveform encoder, a BERT encoder, and a projection layer all contribute to the architecture.
After initial training, the model is refined in ASR by loss of connectionist temporal classification (CTC) to prove it is adaptive in different ways. This is shown in Fig 2.

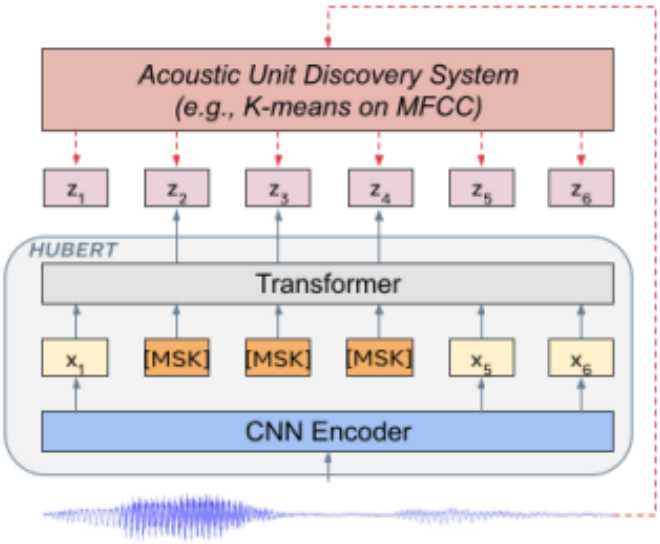


Fig 2. HuBERT Framework

C. DISTILHUBERT

This paper introduces DistilHuBERT, a new multitasking learning algorithm designed to extract hidden signals directly from the HuBERT model[7]. The motivation behind DistilHuBERT comes from computational complexity due to large memory requirements and high training costs associated with state-of-the-art self-supervised language level learning methods such as wav2vec 2.0 and HuBERT. DistilHuBERT uses a teacher-student framework to teach discourse level through multitasking knowledge distillation. Traditional knowledge extraction methods teach students to achieve teacher development, but self-monitored speech samples contain different information at different levels, such as speaker identity or interpretation. To address this DistilHuBERT uses a method in which students' hidden positions are learned from a particular teacher position. For this, the multitasking learning algorithm encourages the converter encoder to create compact representations for many prediction vertices. DistilHuBERT objective functions include minimizing the distance and maximizing the cosine similarity between teacher and student latent representations. This objective ensures that the learner learns to create latent representations of multiple learners from shared positions. The number of prediction heads can be adjusted considering that copies of adjacent lines may contain

the same information. In practice, only a certain number is predicted, reducing computational costs. DistilHuBERT starts with the first two HuBERT CNN extractor and transformer layers. The model is then optimized through prior training, the parameters are frozen and used in downstream applications. The proposed scheme aims to make advanced self-supervised lecture-level learning accessible to a wider audience by dramatically reducing computer requirements. The detailed architecture is shown in Fig 3.

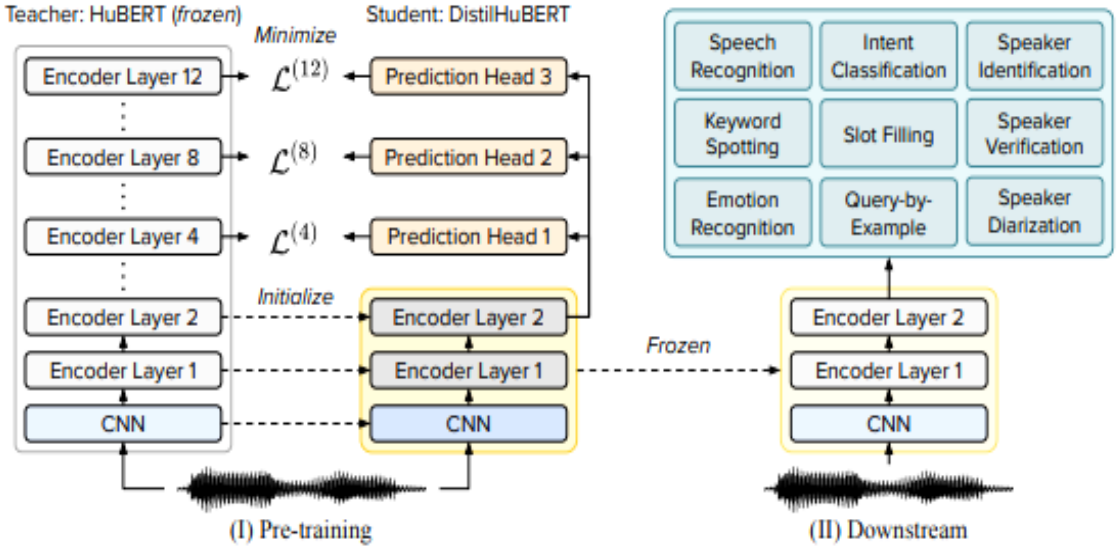


Fig 3. DISTILHuBERT Architecture

D. TEXTLESS TRANSLATOTRON

This work addresses the challenge of language-to-language translation (S2ST), focusing on the limitations of existing end-to-end systems that rely heavily on text management during training, making it unsuitable for languages that do not have a written form, which aims to train a direct end-to-end S2ST model without any textual checks[8]. Key features of the Textless Translatotron include speech spelling, speech synthesis, and sound effects. Notably, multilingual mechanisms based on Vector Quantized Variational Autoencoders (VQ-VAE) have been introduced to extract discrete representations from the target language, guiding linguistic decoder training. This variation deals with missing data.The textless translatotron, designed for end-to-end speech-to-speech translation without text control, includes a speech encoder starting from a pre-separated w2v-BERT model, and a speech quantizer for consumption role for Vector Quantized Variational Autoencoders (VQ-VAE) extraction of discrete speech signals. Two types of quantizers, such as random quantizers and learned quantizers (linear and transformer quantizers) address the missing data by pretraining the speech data of the target language. Linguistic decoders direct training through quantizers on the use of distinctive images rather than text control. Simplified sound effects comparable to Translatotron 2 include distractors modified to match speech signal and spectrogram length. The basic framework of architecture is shown in Fig 4. Textless Translatotrons, competitive performance is tested on both multilingual and bilingual datasets, outperforming previous textless models by +18.5 BLEU on the Fisher Spanish-English corpus.

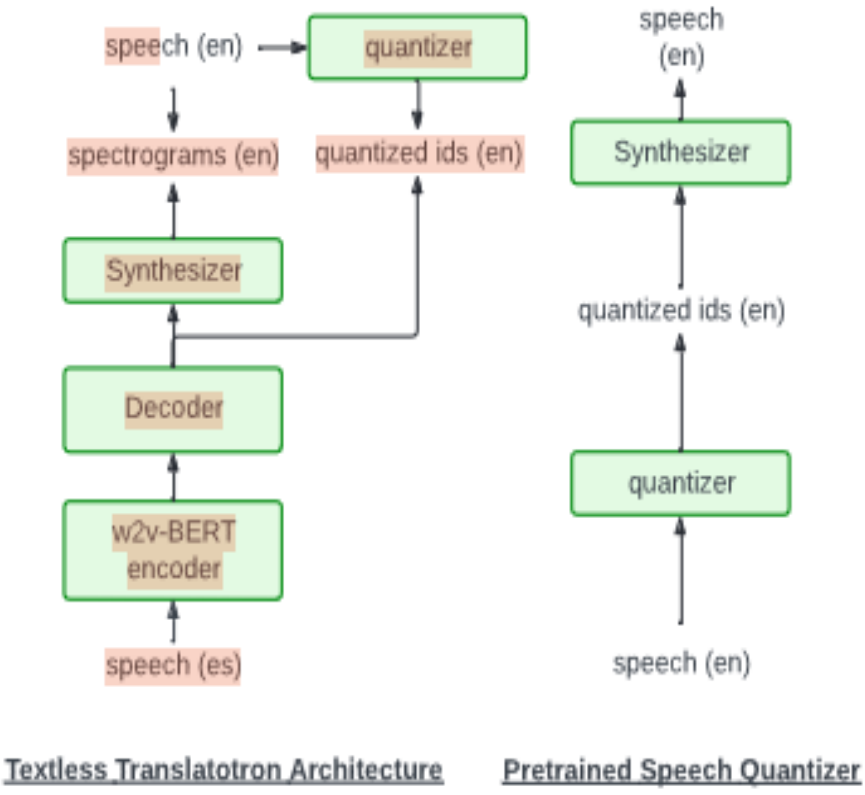


Fig 4. Framework for Textless Translatotron

III. PROPOSED SYSTEM

The proposed system converts spoken words in one language to another directly. It uses pre-trained models and advanced techniques for processing the spoken language input and generating output. The user inputs the audio they want translated into the system.When you feed voice input into this model, it initially uses a self-supervised algorithm to convert the speech into continuous representations. The S2UT model refines this by predicting discrete units using a transformer-based approach. Multitask learning incorporates additional tasks during training, enhancing the model's understanding. For written languages, a CTC decoding process aids in dual modality output. Then the unit-based vocoder converts the discrete units into a waveform.

This process involves training the vocoder for stacked and reduced outputs, with the latter incorporating a lightweight duration prediction module. In essence, the model transforms spoken input into a sequence of unique units, representing the corresponding output waveform.This approach ensures that the translation from one language to another occurs as efficiently as possible as there are no cascaded systems that provide opportunities for failures to be amplified. The unified process makes the translation much more efficient than when compared to traditional models.

A. Architecture

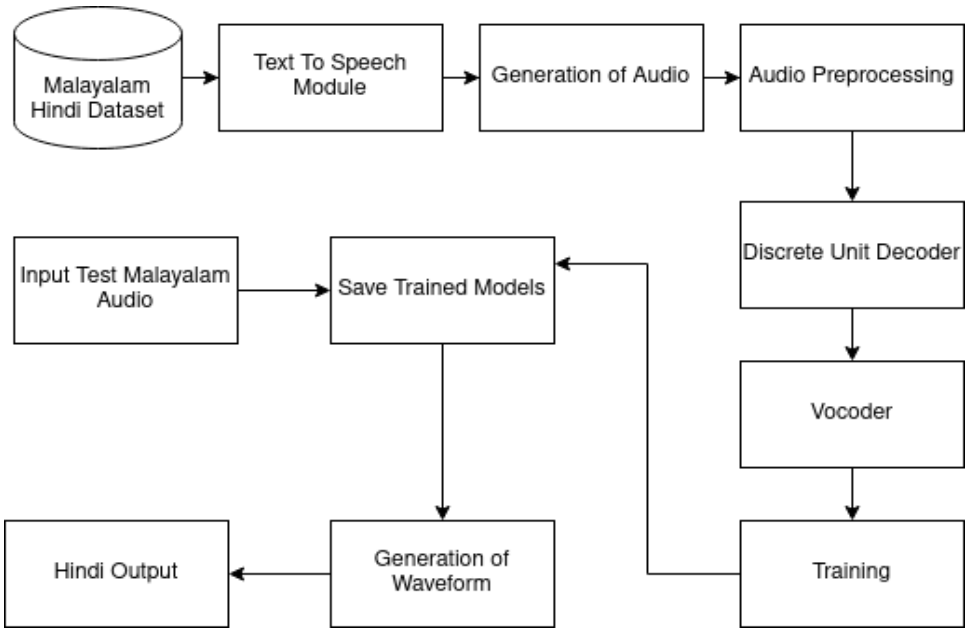


Fig 5. Architecture Framework – Proposed Model

As illustrated in Fig 5, the direct speech-to-speech translation model works by first taking in the input audio and generating from it, discrete units. These units are further sent into the unit vocoder which synthesizes output in the form of waveforms, i.e, the output is directly generated as speech without relying on text generation and text-to-speech techniques.

B. Implementation

The proposed system consists of a transformer-based sequence-to-sequence model with a speech encoder and a discrete unit decoder. Includes supporting functions for advanced learning. The speech-to-group translation (S2UT) paradigm uses self-supervised speech representation learning, which transforms the input language into a continuous representation. Based on the transformer model for machine translation, the S2UT model uses a 1D convolutional layer for down sampling and uses both "stacked" and "reduced" methods to predict discrete units, providing practical training and inference successfully. Multitask learning involves tasks that aid in training, such as CTC decoding of target text for dual modality output. The unit-based vocoder uses HiFi-GAN to convert discrete units into waveforms. For stacked output, the vocoder is trained on discrete units only. Data Augmentation techniques such as mean variance, SpecAugment is used during audio processing in the training phase for higher accuracy.

IV. COMPARATIVE STUDY

In Table 1, the methods used in each paper are listed and compared with each other. Each method is found to have different strengths and weaknesses, with some methods covering for flaws in other methods. This is explicitly shown as the DISTILHuBERT model addresses computational efficiency and speed of the HuBERT model by possibly trading off a degree of accuracy provided by the latter.

Table 1 Comparing Methods

Paper	Methods	Advantages	Disadvantages
Direct Speech-to-Speech Translation With Discrete Units	Discrete Units, S2UT	It can generate consistent speech and text output in one inference pass, which is useful for applications that require both modalities.	It relies on synthetic target speech for model training, which may not reflect the naturalness and diversity of real speech.
HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units	HuBERT	State-of-the-Art Results, Effective use of low quality clusters	Reliance on noisy labels, Simple k means clustering
DistilHuBERT: Speech Representation Learning By Layer-wise Distillation Of Hidden-unit Bert	DISTILHuBERT	Small and fast: DistilHuBERT reduces HuBERT's size by 75% and speeds up inference by 73%	Bad trade-off between the model size and performance.
Textless Direct Speech-to-Speech Translation with Discrete Speech Representation	Textless Translatotron	Can work with languages that don't have a written form or have different regional styles more easily	Data Scarcity, Lack of Textual Supervision, Grapheme-to-Phoneme Conversion

Furthermore, while the Textless Translatotron is designed to primarily work with languages lacking a written form or with diverse regional styles, S2UT is advantageous for applications requiring both speech and text output in one pass. The former's data scarcity challenges could also possibly be challenged by using DISTILHuBERT as it allows faster experimentation and adaptation to different languages.

V. CONCLUSION

In a dynamic language-to-language translation environment, powerful solutions emerge to meet the growing need to extract meaningful insights from comments as the volume of comments increases and their implementation faces challenges such as information overload, time constraints, and course preferences. Carefully designed to integrate different parts of text, this model provides a user-centered solution to these challenges. By using advanced speech recognition and translation techniques, the model empowers users to provide concise and meaningful summaries of linguistic data. Its flexibility, whether handling text or audio extraction, highlights its versatility to meet the user's needs. This research is particularly helpful in increasing the availability and usefulness of the claims in various industries. The model not only simplifies the experience of searching for feedback but also helps develop personalization and effective communication. Looking forward, the model lays the groundwork for further advancements in speech-to-speech translation technology, emphasizing efficiency, accessibility, and adaptability. As the realm of spoken content continues to evolve, the model showcases the potential of integrating speech information to redefine user interactions and extract value from spoken content.

VI. REFERENCES

[1] Takatomo Kano, Sakriani Sakti, Satoshi Nakamura., "End-to-End Speech Translation with Transcoding by Multi-Task Learning for Distant Language Pairs", IEEE/ACM Transaction on Audio, Speech, and Language Processing, Vol. 28, 2020

[2] Sanyuan Chen, Yu Wu, Zhuo Chen, Jian Wu, Jinyu Li, Takuya Yoshioka, Chengyi Wang, Shujie Liu, Ming Zhou., "Continuous Speech Separation with Conformer", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021

[3] Shun-Po Chuang, Alexander H. Liu, Tzu-Wei Sung, Hung-yi Lee., "Improving Automatic Speech Recognition and Speech Translation via Word Embedding Prediction." IEEE/ACM Transactions on Audio, Speech, and Language Processing (Volume: 29), 2020

[4] Xuxin Cheng, Qianqian Dong, Fengpeng Yue; Tom Ko, Mingxuan Wang, Yuexian Zou., "M3ST: Mix at Three Levels for Speech Translation", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023

[5] Ann Lee, Peng-Jen Chen, Changhan Wang, Jiatao Gu, Sravya Popuri,"Direct Speech-to-Speech Translation with Discrete Units" Meta AI, Johns Hopkins University, arXiv:2107.05604v2 [cs.CL] 21 Mar 2022.

[6] Wei-Ning Hsu, Benjamin Bolte, Yao-Hung Hubert Tsai, Kushal Lakhotia,"HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units", arXiv:2106.07447v1 [cs.CL] 14 Jun 2021.

[7] Heng-Jui Chang, Shu-wen Yang, Hung-yi Lee, "DistilHuBERT: Speech Representation Learning by Layer-wise Distillation of Hidden-unit Bert", arXiv:2110.01900v4 [cs.CL] 28 Apr 2022.

[8] Xinjian Li, Ye Jia, Chung-Cheng Chiu, Google Research, Carnegie Mellon University, Tomato AI, "Textless Direct Speech-to-Speech Translation with Discrete Speech Representation" , arXiv:2211.00115v1 [cs.CL] 31 Oct 2022.

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

A Novel Approach for Eye Disease Classification Using Transfer Learning

Neenu Daniel[1], Abhaya Pathrose[2], Eeva E P[2], Gopika Shine[2], Athulya Rose Tom[2]
[1]Assistant Professor, Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Vazhakulam, Kerala, India
[2]Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Vazhakulam, Kerala, India

ABSTRACT
Over the past few years, the field of computer vision and deep learning has experienced an unparalleled expansion, attributed to the exponential growth of computational infrastructure. This trend is evident in the realm of retinal image analysis, where successful artificial intelligence models have been crafted for diagnosing various retinal diseases. This study offers a comprehensive examination of recent deep learning strategies employed in diagnosing five major eye conditions: diabetic retinopathy, glaucoma, age-related macular degeneration, cataract, and myopia. The primary goal of this research is to create and evaluate different transfer learning models capable of accurately diagnosing eye diseases from retinal images. This research aims to provide valuable insights into the advancements, challenges, and methodologies in the field of automated eye disease classification, empowering healthcare practitioners and stakeholders with knowledge for informed decision-making and the promotion of sustainable practices in ophthalmic care.

Keywords: CNN, Deep Learning, Transfer Learning.

I. INTRODUCTION

As stated in the report of the World Health Organization in 2019, it is estimated that there are 1.2 billion individuals with an eye condition. The intricacies of human vision rely heavily on the intricate functions of the eyes. Insufficient focus on eye diseases arises from disparities in [1] limited integration of health services, treatment, and rehabilitation; subpar quality in preventive services; and inadequate health coverage. In [2] the domain of ophthalmology, the use of techniques in deep learning has proven to be impactful in the particularly early and accurate diagnosis of ocular conditions. The retina, derived from the neuroectoderm, is a delicate and transparent sheet of tissue positioned at the back of the eye, playing a crucial role in light sensing. One non-invasive diagnostic test, known as retinography, serves to identify various eye conditions, such as glaucoma, diabetic retinopathy, cataract, hypertensive retinopathy, and diabetic macular edema. Retinographies provide visual insights into the examined structures of the fundus, such as the nearly transparent retinal parenchyma,

displaying a reddish colour attributed to the Retinal Pigment Epithelium (RPE), which may vary based on factors like race, age, and skin colour. Over time, the retina loses its shine.

This study focuses on classification of common eye diseases namely glaucoma, cataract, age-related macular degeneration, diabetic retinopathy, healthy eyes and pathological myopia. Cataract is a common eye condition characterized by clouding of the natural lens within the eye, leading to gradual loss of vision. Cataracts are often age-related, with the majority of cases occurring in individuals over 60 years old. Glaucoma is a group of eye diseases characterized by [3] damage to the optic nerve that is crucial for transmitting visual information from eye to brain. In [4] people with diabetes can develop sight problems due to damage to the light-sensitive tissue in the back of their eyes, called diabetic retinopathy. Age-related macular degeneration is a gradually advancing ocular condition primarily impacting the macula, the central region of the retina. Pathological myopia also known as high myopia is a severe form of near-sightedness characterized by an elongated eyeball and an increased risk of associated eye complications.

The application of deep learning techniques, particularly Convolutional Neural Network has [5] significantly advanced in medical image classification, including the diagnosis of eye diseases. The selected diseases [6], including cataract, glaucoma, diabetic retinopathy, age-related macular degeneration and pathological myopia represent a diverse range of ocular conditions with [7] distinct features. Integrating transfer learning into our classification not only allows us to capitalize on the knowledge encoded in large-scale datasets but also enables the model to learn relevant features crucial for distinguishing subtle differences between these diseases.

The subsequent sections are organized as follows. Section II tells about CNN structures used for the detection. Section III discusses a novel scheme for diagnosis of eye diseases. Section IV presents comparative study on various diseases.

II. LITERATURE SURVEY

A. Deep symmetric convolutional neural network

The study described in [1] introduces the utilization of Deep Symmetric Convolution Structure technology for the detection of two distinct types of lesions associated with Diabetic Retinopathy, a condition that can potentially lead to blindness in diabetic individuals. The primary objective is to enhance the models' capability to identify and categorize these targets effectively. The investigation employs the DIARETDB1 public database, comprising 89 digital fundus images with dimensions of 1500*1512 pixels. In this strategy, a 1*1 convolutional kernel is deliberately selected for the architecture to boost non-linearity and reduce parameters, with the ultimate goal of improving the model's generalization performance, as depicted in Figure 1. The proposed approach employs diverse network structures for efficient detection of two types of lesions outlined in [10]. During feature extraction, various structures are used to filter information, influencing lesion detection differently. Pooling layers are crucial for extracting key information. During the pre-processing phase, the fundus image undergoes segmentation into its red, green, and blue channels. The feature filtering module utilizes specific layers like convolutional, max-pooling, and average-pooling to extract lesion features, enhancing overall method efficacy.

The outcomes of the experiments reveal that the network model incorporating pooling layers achieves higher accuracy compared to the model relying on convolutional layers. The detection accuracy for objects reaches 92.0%, when employing various filtering structures, thereby enhancing both detection performance and feature

extraction. It's worth noting that the evaluation was conducted solely on a single dataset, DIARETDB1, potentially constraining the applicability of the findings to other datasets. Additionally, the method entails a sophisticated network structure, demanding considerable computational resources and time for both training and testing processes.

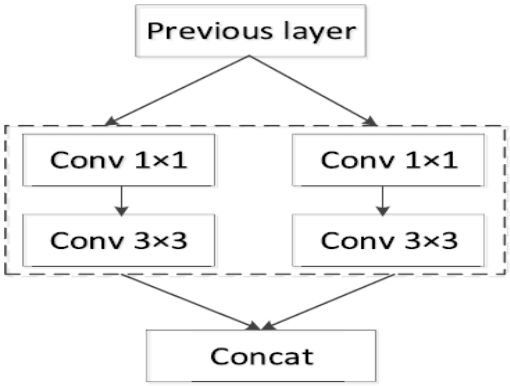


Figure.1 Symmetric Convolution structure

B. Convolutional neural network

In the framework presented in this study [11], the cornerstone lies in the utilization of Convolutional Neural Networks as the fundamental deep learning architecture. The CNN structure comprises a total of six layers, with the initial four layers being convolutional and the last two layers being fully connected. The result from the ultimate fully connected layer is fed into an advanced classifier to identify the presence of glaucoma. Notably, our proposed architecture incorporates response-normalization layers along with overlapping layers to enhance the learning process. The model is trained using a dataset consisting of large number of images from the Optic Nerve Head Research and Glaucoma (ORIGA) database for training purposes, while the evaluation is conducted on all 1678 images from the Chinese Eye Study (SCES). The Area Under the Curve (AUC) metric, employed for performance evaluation, yields values of 0.831 for the ORIGA and SCES datasets, respectively. These results underscore the efficacy of our approach in accurately detecting glaucoma, showcasing promising outcomes in both training and testing scenarios. On CNN, merging layers summarize feature statistics in the image area.

The composite layer is composed with a grid of constituent parts isolated by p pixels, all of which reduces the size of the region centered on the merging unit. In this study, we employed a ROI picture like an intrinsic element of CNN in our study, which yields the first compact image which takes less time to analyse than disassembling a disc and a cup. Data augmentation is accomplished during training by dragging 224 * 224 random packs on 256 * 256 pictures, encompassing their horizontal displays, while coaching our networks on such handouts. During the experiment, CNN forecasts its release of five 224 * 224 leaflets, comprising four corner parts as well as a center area, along with their lateral visibility and a predictor of the network's soft layer on these ten leaflets. This test approach will be referred to as a multi-view test (MVT). The proposed deep learning architecture aims to accurately identify glaucoma-related characteristics in ocular images. The study's findings and proposed deep learning framework may be limited in their generalizability to diverse populations or imaging datasets.

C. Classification of Ocular Conditions in Fundus Images

This study [12] introduces a Convolutional Neural Network designed to categorize two specific conditions: Glaucoma and Diabetic Retinopathy. Dimensionality of glaucoma is 3072 x 2048 pixels, and is 700 x 605 that of diabetic retinopathy. For maintaining standard dimension was resized to 250 x 250. In this proposal, a mean filter is applied to remove noisy pixels. The value of each pixel is replaced with the mean of the values from nearby pixels, accomplished through convolution using a designated mask.

In the convolution layers, feature maps are produced by convolving an image with filters whose weights are determined by size. After generating feature maps, the Rectified Linear Unit function is utilized to introduce non-linear characteristics, emphasizing significant features. ReLU is a commonly favoured feature extraction function in this context, as it identifies negative values, setting them to 0, while preserving the original values of positive ones. Moving on to the pooling layers, these layers process the generated maps by either averaging the values or selecting the maximum value among neighbouring pixels. Max-pooling is a widely utilized function in this layer, involving the selection of the highest value within a square pixel neighbourhood. The size of this neighbourhood is pivotal and is typically defined as a square. The convolution and pooling layers can be iterated depending on the intricacy of the dataset, leading to decreased information flow through the following layers. The proposed architecture is shown in Figure.2. During the last phase, the acquired parameters are transformed into a vector, serving as input for the Multilayer Perceptron. Following this, the Convolutional Neural Network undergoes training using these parameters and weights, utilizing functions like softmax or sigmoid in both feed-forward and backpropagation processes. Deep learning is a subset of artificial intelligence methodologies that utilize artificial neural networks, drawing inspiration from the biological structure of the human brain in terms of learning techniques. In the process of deep learning, mathematical representations are automatically acquired to grasp the latent and intrinsic relationships within the input data. The intelligent pattern classifier was assessed using the ORIGA dataset and DIARETDB0 with images in jpg format. The result of accuracy percentage is 99.89%. Observations were made on the five iterations of confusion matrices and are validated through the K-Fold Cross validation. In the scope of this study, the analysis is limited to only two specific eye conditions, with a notable omission of the examination of healthy eye images. This exclusion of healthy images from consideration is identified as a drawback within the research, presenting a potential limitation to the comprehensiveness of the findings.

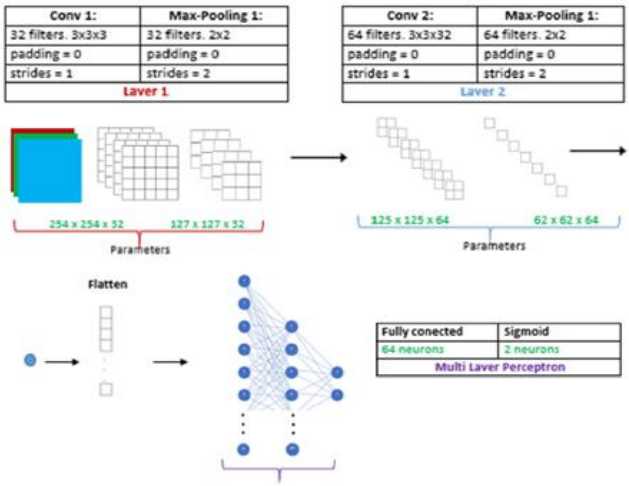


Figure.2 Architecture of the proposed system

D. Mixed models for diabetic retinopathy detection and classification

In reference [13], the researchers utilized the Indian Diabetic Retinopathy Image Dataset (IDRiD) to validate their proposed approach for the detection and classification of diabetic retinopathy (DR). The dataset consists of 516 fundus images, each with a 50-degree view and a resolution of 4288x2848 pixels. It encompasses images depicting different stages of DR, including proliferative DR, normal, moderate, severe, and mild. The category of DR is found out through a hybrid approach involving machine learning algorithms and a majority voting system. The overall process can be outlined as follows: Extracting pertinent features, encompassing colour, texture, and shape features, from pre-processed fundus images is accomplished using methodologies such as histogram of oriented gradients (HOG), grey-level co-occurrence matrix (GLCM), and local binary pattern (LBP). The features extracted are fed into machine learning classifiers, including support vector machine (SVM), binary trees (BT) and K-nearest neighbour (KNN). These classifiers are trained to classify features into distinct classes representing normal, mild, and severe stages of DR. The outcomes of each classifier are amalgamated using a majority voting method as in Figure 3. This entails consolidating the individual classification results from SVM, KNN, and BT classifiers and selecting the class label that receives the maximum number of votes as the conclusive classification for a given fundus image. By integrating multiple classifiers and incorporating a majority voting mechanism, the methodology seeks to fortify the robustness and precision of the DR classification process. This strategy leverages the unique strengths of different classifiers to enhance the overall dependability of disease classification, ultimately contributing to more effective diagnoses and treatments for DR. The hybrid model, which integrates the outcomes of individual classifiers through a majority voting mechanism, demonstrates superior performance compared to standalone models. Specifically, BT model more efficient than SVM, while the KNN is less efficient compared to all other classifiers. Furthermore, the research highlights that heightened disease severity is associated with improved classifier performance. As the severity of DR increases, there is a consistent enhancement in classification accuracy, particularly evident in the more severe stages of the disease.

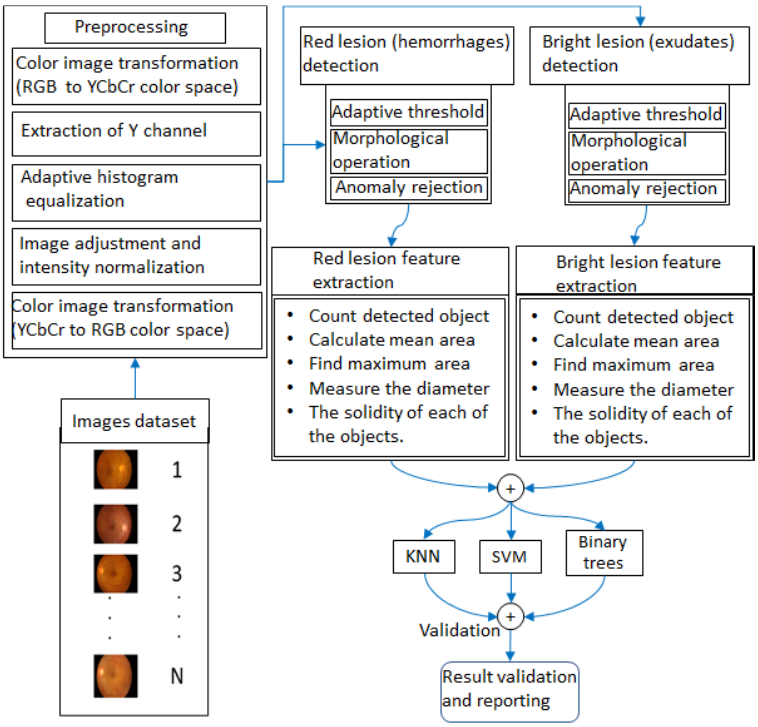


Figure.3 Block Diagram of the Proposed Approach

The primary advantage of utilizing the technology outlined is its capacity to offer a thorough and precise approach to detecting and classifying DR. Through the amalgamation of image processing methods for extracting features for classification, the methodology provides a robust and automated solution for diagnosing DR across various disease severity levels. The incorporation of multiple machine learning classifiers, including SVM, KNN, and BT, along with a majority voting system, improves the precision and dependability of disease diagnosis. Moreover, the methodology's successful performance on the IDRiD demonstrates elevated levels of accuracy, sensitivity, and specificity, underscoring its potential for practical use in real-world clinical scenarios. The technology presented has certain limitations that warrant consideration. One notable disadvantage is the reliance on a single dataset IDRiD, which may limit the generalizability of the methodology to diverse populations and datasets. One of the widely used method. Additionally, the use of a limited feature set for disease detection and classification could potentially overlook important diagnostic markers, impacting the methodology's overall accuracy and reliability. Furthermore, the methodology's evaluation against a restricted set of conventional methods and the use of a single performance metric may limit the depth of insights gained from the comparison and evaluation. These limitations underscore the need for further research and validation to ensure the robustness and applicability of the technology across different populations.

III. PROPOSED SYSTEM

The proposed work aims to offer a comprehensive examination of recent advancements in deep learning (DL) strategies applied to the diagnosis of retinal diseases through the analysis of fundus images. The study concentrates on the classification of six distinct eye conditions: cataract, glaucoma, pathological myopia, diabetic retinopathy, age-related macular degeneration, and healthy eyes. In this context, pretrained transfer learning models, including AlexNet, VGG19, ResNet, and EfficientNetV2, are employed to implement the diagnostic system. It derives its structure and functionality from the human visual system, which can rapidly and precisely recognize objects and patterns in visual data. By obtaining the insights garnered from these established models, the study aims to improve the precision and effectiveness of diagnosing retinal diseases, ultimately contributing to the enhancement of ophthalmic healthcare practices. This study also aims to delineate potential future avenues for emerging researchers intrigued by AI-driven diagnosis of retinal diseases.

A. Architecture

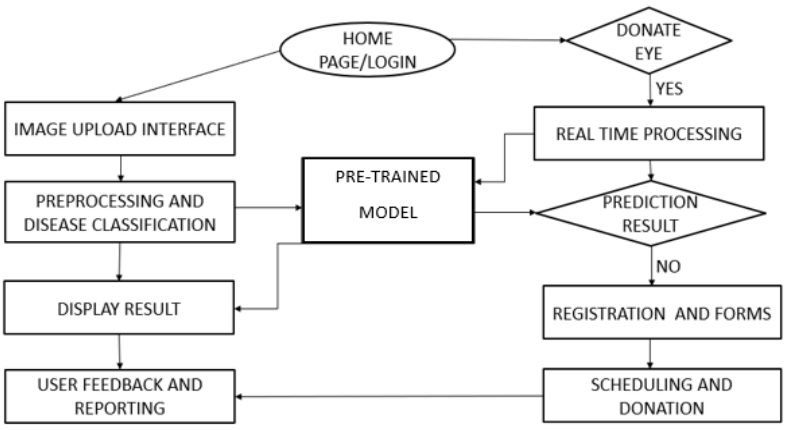


Figure.4 Architecture diagram of proposed system

The system as shown in Figure.4 utilizes deep learning techniques like transfer learning to automatically detect and classify prevalent eye conditions. The user starts at the HOME page/LOGIN, then has the option to donate their eyes or directly proceed to upload an eye image, once the image is uploaded, it is then pre-processed and sent to a pre-trained model for disease classification. This model tries to identify any of the proposed eye diseases in the image. Displays the result, and collects user feedback. If they opt for eye donation, it processes in real-time. If the model predicts that there is no disease in the image, the user is shown the results and asked to register and fill out some forms. This information will be used to connect the user with a potential donor. The system also has a feedback loop where users can provide feedback on the results. This feedback can be used to improve the accuracy of the model over time.

B. Implementation

For the model architecture we use four transfer learning models. Fundus photography involves capturing a two-dimensional image of the 3D retina by projecting reflected light onto an image plane. In our experiment we used Ocular Disease recognition dataset and ODIR5K_classification dataset. Each eye condition consists of more than 400 images. Various performance evaluation metrics are employed to assess the Deep Learning model's effectiveness in tasks related to diagnosing retinal diseases. Transfer learning models undergo training and testing, with the selection of the optimal model determining the final outcome.

IV. COMPARATIVE STUDY

As shown in Table 1 four deep learning approaches are used for eye disease detection, each with its own vibrant hues and subtle shadows.

The DSCNN shines brightly in diabetic retinopathy detection, but its narrow field of view and demanding processing requirements cast long shadows. The glaucoma-hunting CNN excels in laser-sharp precision, but its focus fades when other ailments appear. The CNN classifier stands resolute with k-fold validation, yet its canvas paints only healthy and diabetic retinas. Finally, the mixed-model detective champions versatility and agility, but its path remains shrouded in limited validation and data.

Choosing the best depends on the quest: pinpoint accuracy for a specific foe, a broader map for diverse landscapes, or swift, automated exploration. Remember, the optimal path lies in understanding the needs and harnessing the right light. It is crucial to recognize that the optimal path in selecting a model lies in a comprehensive understanding of the researcher's needs and the ability to harness the right kind of illumination for the task at hand. As medical image analysis continues to advance, the importance of aligning the strengths of these models recognizing the necessity of aligning these models' capabilities with the specific needs of the research becomes increasingly evident.

TABLE I COMPARATIVE STUDY OF LITERATURE SURVEY

SI NO	TITLE	TECHNOLOGY USED	ADVANTAGES	DISADVANTAGES
1.	"A NOVEL DIABETIC RETINOPATHY DETECTION APPROACH BASED ON DEEP SYMMETRIC CONVOLUTIONAL NEURAL NETWORK"	Deep Symmetric Convolutional Neural Network	Enhanced Detection and Performance	Limited Dataset, Computational Complexity
2.	"DETECTION OF GLAUCOMA USING DEEP LEARNING"	Convolutional Neural Network (CNN)	Increase accuracy of glaucoma by using six layers of CNN	Only for detecting glaucoma not for any other eye disease
3.	"CLASSIFICATION OF EYE DISEASES IN FUNDUS IMAGES"	Convolutional Neural Network	The Algorithm has been validated through the K-Fold Cross Validation test, which helps to ensure the reliability of the results	Worked on only two eye conditions, Healthy Images were not analyzed
4.	"DIABETIC RETINOPATHY DETECTION AND CLASSIFICATION USING MIXED MODELS FOR A DISEASE GRADING DATABASE"	Mixed Models (KNN, SVM, Binary Tree)	Improved Accuracy, Generalizability, Efficient and Automated	Limited Validation, Limited Dataset, Limited Feature Set

V. CONCLUSION

The urgency for automated systems to identify eye-related diseases is evident due to the insufficient number of medical experts in comparison to the patient volume. The availability of a colour retinal image, containing a diverse range of ocular-related pathologies in image format, has opened up significant opportunities for research in the realm of medical image analysis. In this study an attempt is made to discuss various transfer learning models. It is comparing AlexNet, VGGNet19, Resnet-50 and EfficientNetV2 models in addressing various eye conditions through image classification. The detailed explanation of various architecture and the optimization techniques employed highlights the efficiency of the model in practical scenarios, accompanied by an acknowledgment of the challenge posed by the need for a substantial volume of labelled data during the training phase.

VI. REFERENCES

[1] T Liu et al., "A Novel Diabetic Retinopathy Detection Approach Based on Deep Symmetric Convolutional Neural Network," in IEEE Access, vol. 9, pp. 1600552-160558, 2021, doi: 10.1109/ACCESS.2021. 3131630..

[2] Ravi Kumar Gupta et al.," Detection of Glaucoma using Deep Learning" International Journal of Research in Engineering, Science and Management Volume 5, Issue 5, May 2022 https://www.ijresm.com | ISSN (Online): 2581-5792

[3] O. Bernabé, E. Acevedo, A. Acevedo, R. Carreño and S. Gómez, "Classification of Eye Diseases in Fundus Images," in IEEE Access, vol. 9, pp. 101267-101276, 2021, doi: 10.1109/ACCESS.2021.3094649.

[4] A. Bilal, G. Sun, Y. Li, S. Mazhar and A. Q. Khan, "Diabetic Retinopathy Detection and Classification Using Mixed Models for a

[5] Disease Grading Database," in IEEE Access, vol. 9, pp. 23544-23553, 2021, doi: 10.1109/ACCESS.2021.3056186.

[6] S. B. Patil and B. P. Patil, Automatic detection of microaneurysms in S. B. Patil and B. P. Patil, ``Automatic detection of microaneurysms in retinal fundus images using modified high boost filtering line detectors and OC-SVM,'' in Proc. Int. Conf. Ind. 4.0 Technol. (I4Tech), Feb. 2020,

[7] Ovreiu, S., Cristescu, I., Balta, F., Sultana, A. And Ovreiu, E., 2020, June. Residual networks early warning of glaucoma. In 2020 13th International Conference on Communications (COMM) (pp. 161-164). IEEE.

[8] A. Colomer, J. Igual, and V. Naranjo, ``Detection of early signs of diabetic retinopathy based on textural and morphological information in fundus images,'' Sensors, vol. 20, no. 4, p. 1005, 2020.

[9] N. Patil, V. P. Rao, and N. P. Patil, ``Hybrid CNN assisted computer aided diagnosis system for glaucoma detection and classification: Glaucoma. NetC,'' Int. J. Innov. Technol. Exploring Eng. (IJITEE), vol. 9, pp. 30603072 Mar. 2019.

[10] P. Khojasteh, L. A. P. Júnior, T. Carvalho, E. Rezende, B. Ali Ahmad, J. P. Papa, and D. K. Kumar, ``Exudate detection in fundus images using deeply-learnable features,'' Comput. Biol. Med., vol. 104, pp. 6269, Jan. 2019.

[11] L. Qiao, Y. Zhu, and H. Zhou, ``Diabetic retinopathy detection using prognosis of microaneurysm and early diagnosis system for non-proliferative diabetic retinopathy based on deep learning algorithms,'' IEEE Access, vol. 8, pp. 104292104302, 2020.

[12] M. K. Nath, S. Dandapat, “Techniques of Glaucoma Detection from Colour Fundus Images: A Review”, I. J. Image, Graphics and Signal Processing, 2012, 9, 44-51.

[13] S. Kumar, F. G. Khan, S. Shah, A. Khan, S. Shamshirband, Z. U. Rehman, I. Ahmed Khan, and W. Jadoon, ``A deep learning ensemble approach for diabetic retinopathy detection,'' IEEE Access, vol. 7, pp. 150530150539, 2019.

[14] S. S. Rahim, V. Palade, J. Shuttleworth, and C. Jayne, ``Automatic screening and classification of diabetic retinopathy and maculopathy using fuzzy image processing,'' Brain Informat., vol. 3, no. 4, pp. 249267, Dec. 2016.

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

An Approach for Nutrient Deficiency Detection and Weather Forecasting

Ierin Babu[1], Abin Thomas[2], Alex Sebin[2], Joel George Toine[2], R Jayakrishnan[2]
[1]Assistant Professor, Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Ernakulam, Kerala, India
[2]Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Ernakulam, Kerala, India

ABSTRACT
In the agricultural domain, farmers encounter significant hurdles concerning nutrient insufficiency in crops and the erratic nature of climate patterns. The deficiency of vital nutrients such as nitrogen, phosphorus, and potassium presents a formidable challenge to agricultural output and overall farm productivity. Simultaneously, the unpredictable weather conditions create uncertainties, impacting crop management and harvest outcomes. To tackle these pressing concerns, this comprehensive review delves into diverse models tailored for detecting nutrient deficiencies in plants and those pivotal in weather prediction.
Keywords: CNN, Ensemble Learning, ANN, LSTM

I. INTRODUCTION

In contemporary agricultural practices, the integration of the latest technologies is indispensable for mitigating the diverse challenges encountered by farmers on a global scale. We investigate the intricate realms of nutrient deficiency detection in plants and the precision forecasting of weather, both of which stand as pivotal elements within this technological convergence.

Agricultural success lies in the effective management of nutrients to ensure optimal crop growth and long-term sustainability. For that, exploring a wide array of methodologies, including statistical models, machine learning algorithms, and advanced deep learning techniques, all aimed at enhancing the assessment of nutrient deficiencies is required. By delving into these techniques, it not only provides a robust understanding of their applications but also offers actionable insights for driving forward agricultural practices.

Furthermore, the ability to accurately predict weather patterns is instrumental in empowering farmers to make timely and informed decisions amidst the changing climatic conditions. Through a meticulous examination of numerical weather prediction models, ensemble methods, etc., we could shed light on their transformative potential in agriculture. Beyond mere forecasts, it explores how these advancements contribute to disaster preparedness, climate resilience, and sustainable farming practices.

By synthesizing the latest research and technology in nutrient deficiency detection and weather forecasting, the aim is to serve as a beacon of knowledge for agricultural stakeholders worldwide. Moreover, by elucidating the interconnectedness between these domains, it not only enriches our understanding but also sparks vital discussions on leveraging technology to navigate the complexities of a changing climate and meet the escalating demands of global food production.

II. LITERATURE SURVEY

Rice is a vital crop for more than half of the world's population and plays a crucial role in Bangladesh's agriculture. However, its production faces challenges due to nutrient deficiencies that can significantly impact yields. Detecting these deficiencies early is essential for improving productivity, and researchers are exploring advanced technologies such as computer vision solutions powered by Convolutional Neural Networks (CNNs) and transfer learning from ImageNet. Specifically, the use of the DECNN model, adapted from DenseNet169, has demonstrated impressive accuracy in identifying nutrient deficiencies in rice plants.[6] Transfer learning, a machine learning technique that applies knowledge from one task to enhance performance in another, has proven especially valuable in this context, enabling the model to achieve robust results even with limited data availability.[5]

Additionally, ensemble learning techniques have emerged as powerful tools in machine learning, offering improved predictive performance by combining insights from multiple models.[10] In the research conducted, a weighted ensemble comprising DenseNet169, DenseNet201, and InceptionV3 was employed, giving greater weight to the most effective models. This strategic fusion illustrates the effectiveness of ensemble methods in enhancing overall predictive capabilities.[7] While this approach offers significant advantages, including enhanced model robustness through data augmentation and the integration of multiple pre-trained models, it is not without its challenges. Factors such as increased training time due to the complexity of processing augmented data and the intricacies of hyperparameter tuning highlight the nuanced nature of machine learning in agricultural applications.

Nonetheless, these efforts underscore the promising potential of advanced technology in addressing agricultural challenges and promoting global food security.[1]

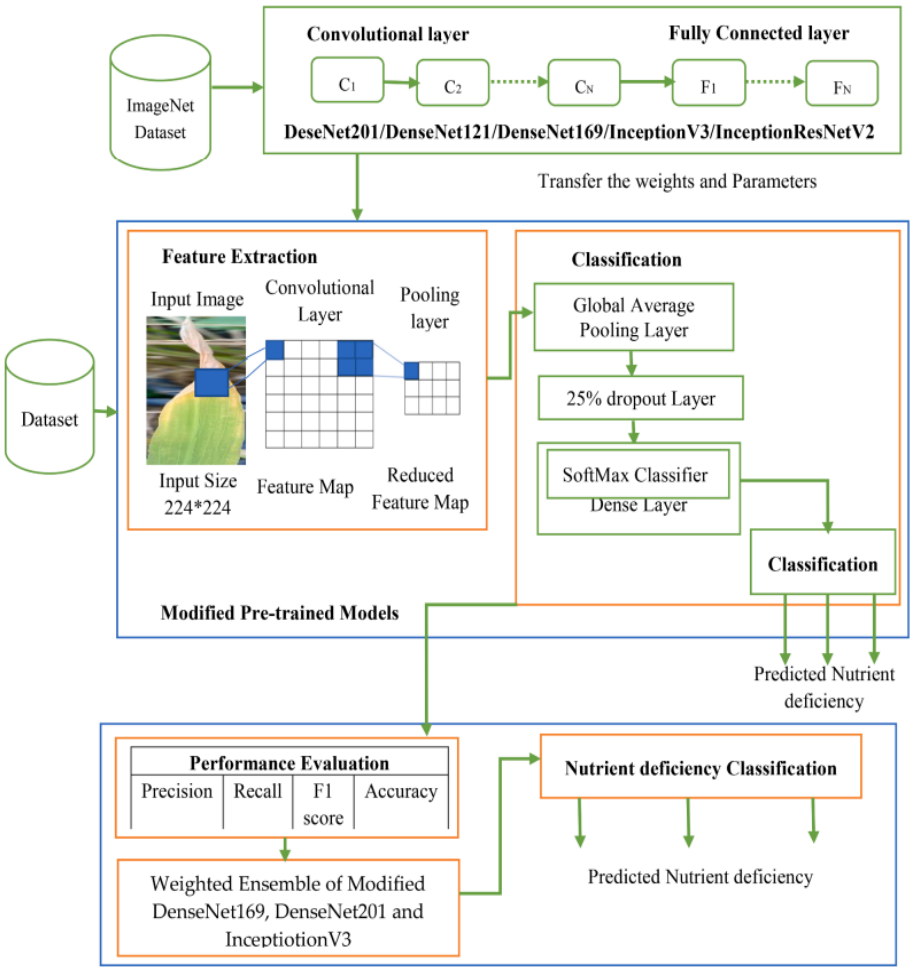


Figure 1: Architecture diagram

The timely identification of nutrient deficiencies is crucial in preventing substantial agricultural losses and enhancing overall yield while promoting environmentally friendly fertilizer usage.[8] This study uses deep neural networks and transfer learning to efficiently recognize nutrient deficiency symptoms in RGB images, aiming to advance early diagnosis and improve agricultural yield sustainability. Employing two real-world datasets—one for sugar beets and another for orange trees—the research achieves remarkable performance with 98.65 percent and 98.52 percent Top-1 accuracies, respectively, using fine-tuning with EfficientNetB4. Despite demonstrating accurate analysis through Grad-CAM++, the study finds that agricultural transfer learning does not contribute to performance improvement. This underscores the effectiveness of the fine-tuned EfficientNetB4 in diverse agricultural scenarios for timely nutrient deficiency identification.

Transfer learning is the technique of applying knowledge gained from one problem domain to a different but related domain. Specifically, the paper discusses the use of pre-trained Convolutional Neural Networks, initially developed for tasks like image recognition on large datasets such as ImageNet and adapting them to identify nutrient deficiencies in chili plants.[9] This approach leverages the generic features learned by the CNNs on a vast array of images and fine-tunes them with a smaller, domain-specific dataset of chili plant images. The result is a model that can accurately diagnose nutrient deficiencies with less data and training time than would be required to train a model from scratch.

The EfficientNetB4 model featured in the research paper stands out as a cutting-edge CNN belonging to the EfficientNet family, renowned for its effectiveness and superior accuracy. It adopts a unique compound scaling technique, which uniformly adjusts the network's depth, width, and resolution based on a compound coefficient. In the context, EfficientNetB4 underwent pre-training on the ImageNet dataset and subsequently underwent fine-tuning using a specific dataset containing images of chili plants to discern nutrient deficiencies. This strategic transfer learning approach enabled the model to achieve exceptional accuracy while requiring less data and computational resources compared to training a model from scratch.[3]

In response to the challenges posed by traditional weather forecasting methods, particularly in the context of climate change, innovative approaches leveraging deep learning techniques like Long Short-Term Memory (LSTM) and Transductive Long Short-Term Memory (T-LSTM) models have gained prominence. The primary focus is on enhancing precision for hydrological variables through machine learning on time series data. The LSTM model, a specialized form of recurrent neural network (RNN), is employed due to its effectiveness in handling spatiotemporal data and addressing issues like the vanishing gradient problem. Key components within the LSTM architecture, including input and forget gates, memory cell updates, and output calculations, enable the model to capture long-range dependencies in data. The T-LSTM model, an iteration of LSTM, further improves performance by incorporating the impact of training data points and considering the similarity between them. Despite its advantages in generalization capability, enhanced learning ability, and consideration of nearby data points, the T-LSTM model faces challenges such as resource-intensive training processes and reduced interpretability due to its complexity.

The LSTM model's methodology relies on its ability to process sequential information efficiently, utilizing elements like the sigmoid activation function, input at a given time, previous hidden and cell states, and element-wise multiplication. On the other hand, the T-LSTM model introduces a sequence-specific subscript value to tailor its processing based on the distinct characteristics of the T-LSTM sequence, aiming to improve overall performance compared to the standard LSTM model. While the T-LSTM model exhibits notable advantages in capturing information from the training data and considering the influence of nearby data points in the test dataset, challenges arise in terms of computational resource requirements and the interpretability of its decision-making process. These considerations highlight the ongoing efforts to balance the advantages and drawbacks of advanced deep learning models for weather forecasting in the face of evolving climatic conditions.[2]

Researchers developed a straightforward water level prediction model named FDPRE, crucial for anticipating farmland waterlogging. The model integrates four machine learning (ML) algorithms and incorporates weather forecasts to accurately predict water levels. Notably, the Random Forest (RF) and Multiple Perception Model (MLP) emerged as top-performing algorithms, achieving impressive R2 values ranging from 0.7180 to 0.9803 and 0.5717 to 0.9965. The study demonstrates that the integration of machine learning techniques, specifically RF and MLP, along with weather forecasts, is an effective method for forecasting pivotal factors driving waterlogging and assessing potential disaster losses in farmland.

Furthermore, the study highlights the economic implications of waterlogging by examining a one-day flooding scenario under different rainfall conditions. Economic losses due to waterlogging were substantially higher under 100 mm rainfall ($23.53 million) compared to 50 mm rainfall ($12.69 million). Additionally, the research indicates that the lower reaches of the Sihu basin experienced a greater reduction in crop yield under both rainfall scenarios, emphasizing the significance of accurate predictions for mitigating the impact of

waterlogging on farmland. The findings underscore the importance of utilizing machine learning models, such as RF and MLP, in conjunction with weather forecasts for effective farmland waterlogging prediction and management.[4]

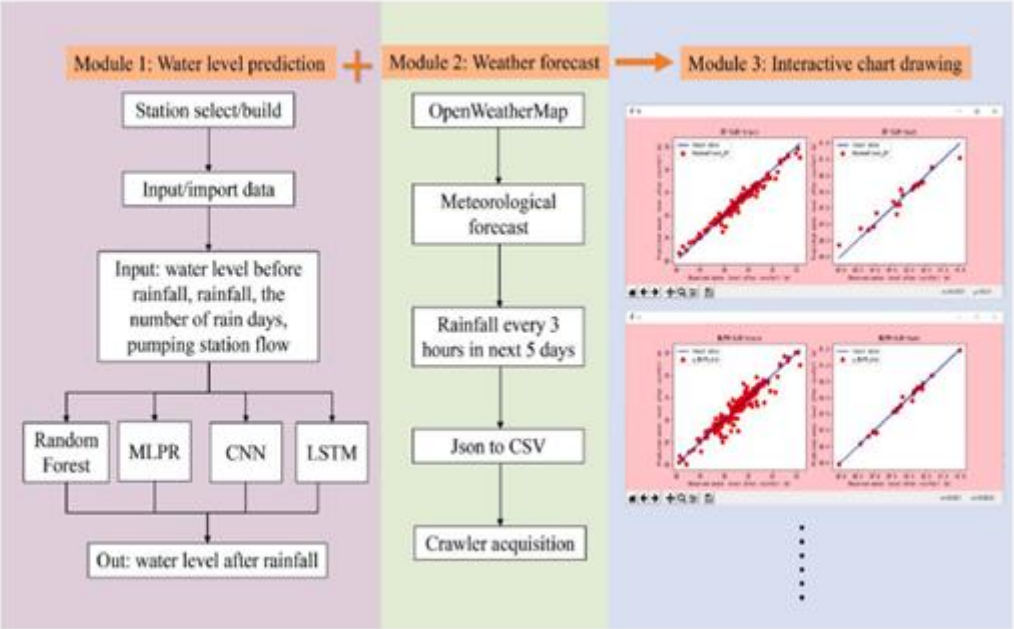


Figure 2: Architecture diagram

III. PROPOSED SYSTEM

The proposed system introduces an innovative approach to crop management by implementing an ecommerce platform for crop buying and selling along with a Convolutional Neural Network (CNN) model for the detection of nutrient deficiencies in rice crops. The input provided for nutrient deficiency detection consists of images of rice crops, constituting the dataset used to train the CNN model. In the process of image analysis, the CNN model can rapidly and accurately identify nutrient imbalances in rice plants. The output of this process includes a detailed analysis report, specifying the types and severity of nutrient deficiencies detected in the crops. Simultaneously, the system also incorporates a regression model for weather forecasting. This involves analyzing historical weather data obtained from the Kaggle dataset, selecting relevant features, and training the regression model to predict future weather conditions. The output from the regression model provides farmers with reliable and timely weather forecasts.

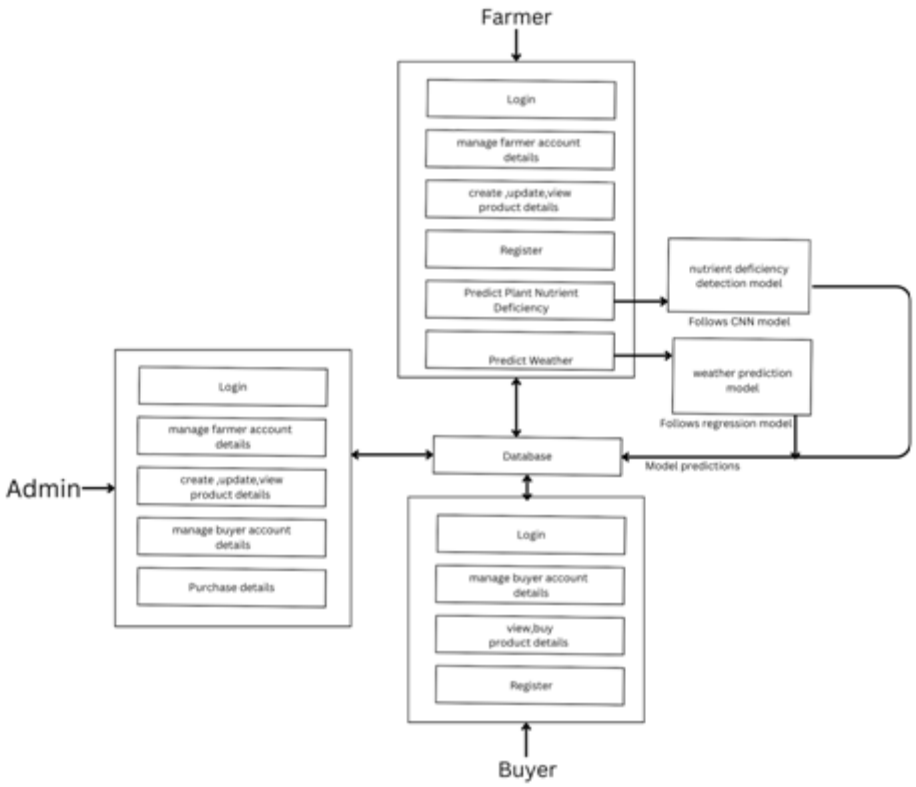


Figure 2: Component Architecture

A. Architecture

The system integrates farmers and buyers, offering them a platform for product transactions, augmented with nutrient deficiency detection and weather forecasting modules. In this setup, inputs for nutrient deficiency detection undergo processing by a Convolutional Neural Network (CNN) model, generating relevant results. Similarly, an Artificial Neural Network (ANN) model receives inputs for weather forecasting, enabling precise predictions of weather conditions.

B. Implementation

Nutrient deficiency detection begins with gathering plant images depicting both deficient and healthy states, ensuring diversity in sourcing and expert labeling. These images undergo standardization, resizing, and pixel normalization during preprocessing, with additional techniques like augmentation and feature extraction applied. The extracted features, indicative of deficiency patterns, feed into a Convolutional Neural Network (CNN) model for training. Through iterative learning, the model refines its ability to discern these patterns accurately. Once trained, it analyzes unseen plant images to predict deficiencies, offering unique insights into their nutritional health.

Weather forecasting begins with the compilation of historical weather data, encompassing temperature, precipitation, humidity, and wind speed. This data undergoes preprocessing, including the handling of missing values, normalization for consistent scales, and addressing outliers to enhance model performance. Subsequently, a regression model is trained on the refined data, iteratively establishing correlations between different weather parameters. Through this training process, the model adjusts its parameters to minimize the variance between predicted and actual weather conditions. When deployed, the trained regression model

utilizes current input data, such as dates and other relevant factors, to generate accurate weather predictions, leveraging its learned relationships to provide reliable forecasts.

All paragraphs must be indented. All paragraphs must be justified, i.e. both left-justified and right-justified.

IV. COMPARATIVE STUDY

The below table shows the comparison between the different models .

Paper	*Model*	*Advantage*	*Disadvantage*
Nutrients deficiency diagnosis of rice crop by weighted average ensemble learning	DECNN	Data augmentation and the utilization of five modified pretrained models, contributing to enhanced model robustness and performance	Complexity of processing augmented data and challenges associated with hyperparameter tuning. Additionally, integrating multiple models may pose logistical challenges
Using EfficientNet and transfer learning for image-based diagnosis of nutrient deficiencies	EfficientNetB4	Grad CAM++ technique to visualize the salient regions in the images and uses deep learning and transfer learning for image-based diagnosis.	Does not address the imbalanced nature of the datasets, which may lead to biased or unstable results.
DWFH: An improved data-driven deep weather forecasting hybrid model using Transductive Long Short-Term Memory	Transductive Long Short Term Memory	The model excels in generalization and learning, effectively considering nearby data points for robust predictions.	The model's resource-intensive training and complex architecture limit efficiency and interpretability, posing challenges in practical applications.
Coupling machine learning and weather forecast to predict farmland flood disaster: A case study in Yangtze River basin	Random Forest approach	Improved flood prediction and integration of weather forecasts.	Complexity of deep learning models , Huge data dependency and lack of essential data.

V. CONCLUSION

In conclusion, this review provides a comprehensive analysis of methodologies employed in detecting nutrient deficiencies in plants and forecasting weather patterns. Through an exploration of diverse approaches, it illuminates their efficacy and constraints. The significance of technology in tackling agricultural complexities becomes evident, emphasizing potential synergies between advancements in plant science and meteorology. Amidst the uncertainties of climate change and escalating food demands, these insights provide valuable guidance for developing innovative and sustainable farming practices.

Advanced techniques like weighted average ensemble learning, EfficientNet with transfer learning, and Transductive Long Short-Term Memory show promise in facilitating quicker generation of results for agricultural applications, but they entail complexities and resource-intensive requirements. In contrast, simpler models like normal CNNs and regression models offer ease of interpretation and potentially lower resource needs while still providing valuable insights.

Future implementations should focus on optimizing these simpler models by addressing data preprocessing challenges, exploring feature engineering opportunities, fine-tuning algorithms for efficiency, integrating diverse data sources, and ensuring scalability with larger datasets and computing resources.

VI. REFERENCES

[1] Md. Simul Hasan Talukder and Ajay Krishno Sarkar. "Nutrients deficiency diagnosis of rice crop by weighted average ensemble learning" in Elsevier, August 2023.

[2] K. Venkatachalam, Pavel Trojovsky, Dragan Pamucar, Nebojsa Bacanin, and Vladimir Simic. "DWFH: An improved data-driven deep weather forecasting hybrid model using Transductive Long Short-Term Memory (T-LSTM)" in Elsevier, March 2023.

[3] Borja Espejo-Garcia, Ioannis Malounas, Nikos Mylonas, Aikaterini Kasimati, and Spyros Fountas. "Using EfficientNet and transfer learning for image-based diagnosis of nutrient deficiencies" in Elsevier, March 2022.

[4] Zewei Jiang and Shihong Yang. "Coupling machine learning and weather forecast to predict farmland flood disaster: A case study in Yangtze River basin" in Elsevier, March 2022.

[5] Sagar, A., & Jacob, D. (2021). On using transfer learning for plant disease detection. BioRxiv, 2020-05.

[6] G.S. Khush. What it will take to Feed 5.0 Billion Rice consumers in 2030. Plant Mol. Biol. 59 (2005) 1–6.

[7] Rice consumption by country 2019. Statista. Available online: https://www.statista.com/statistics/255971/top-countries-basedon-rice--consumption-2012-2013/ (accessed on 1 April 2022).

[8] T. Makkar and Yogesh. A Computer Vision Based Comparative Analysis of Dual Nutrients (Boron, Calcium) Deficiency Detection System for Apple Fruit, in: Proceedings of the 4th International Conference on Computing Communication and Automation (ICCCA), 2018, pp. 1-6.

[9] U. Watchareeruetai, P. Noinongyao, C. Wattanapaiboonsuk, P. Khantiviriya, S. Duangsrisai. Identification of plant nutrient deficiencies using convolutional neural networks, in: Proceedings of the International Electrical Engineering Congress (iEECON), 2018, pp. 1–4.

[10] V. Perumal, V. Narayanan, S.J.S. Rajasekar. Detection of COVID-19 using CXR and CT images using transfer learning and haralick features. Appl. Intell. 51 (2021) 341–358.

[11] A. Vulli, P.N. Srinivasu, M.S.K. Sashank, J. Shafi, J. Choi, M.F. Ijaz. Fine-tuned denseNet-169 for breast cancer metastasis prediction using FastAI and 1-cycle policy. Sensors 22 (8) (2022) 2988, https://doi.org/10.3390/s22082988 (Basel) Apr 13PMID: 35458972; PMCID: PMC9025766.

[12] Zewei Jiang and Shihong Yang. "Coupling machine learning and weather forecast to predict farmland flood disaster: A case study in Yangtze River basin" in Elsevier, March 2022.

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

SFLRS: Supervised Feature-Level Rating System

Ms Dona Jose*[1], Anand Vishnu K V[2], Jerin Joseph[2], Krishna Renjith[2], Liju Mon A P[2]
*[1]Assistant Professor, Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Muvattupuzha, Kerala, India
[2]Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Muvattupuzha, Kerala, India

ABSTRACT
The study explores the use of supervised learning techniques to produce feature-level ratings for mobile products from customer reviews and review votes. The primary aim is to enhance decision-making for both consumers and manufacturers. In contrast to traditional product-level ratings, feature-level ratings offer a more detailed comprehension of particular product characteristics, enabling producers to improve their goods and enabling customers to make individualized purchasing choices. Based on mobile reviews from Amazon, the analysis applies feature focused Sentiment Analysis and Supervised Learning to produce ratings for different features. These insights contribute to the advancement of recommender systems, consumer research, and other applications, offering valuable implications for product enhancement and personalized purchasing choices.
Keywords: Natural Language Processing, Supervised Learning, Recommendation System.

I. INTRODUCTION

In the current landscape of burgeoning online shopping, where the Internet plays a pivotal role in shaping consumer habits, the need for a sophisticated product rating system becomes increasingly apparent. Traditional product-level ratings, while widely used, often lack the granularity necessary for informed decision-making, particularly for consumers who prioritize specific product features. Recognizing this gap, the proposed solution is a comprehensive feature-level rating system designed to cater to the diverse preferences of online shoppers. The system capitalizes on customer reviews, sentiment analysis, and review votes to generate nuanced ratings for individual product features. This approach goes beyond generic product-level evaluations, offering a more detailed and tailored assessment.

Building a feature-level rating system involves overcoming several challenges. First and foremost is the identification of relevant features. To address this, a systematic approach employs word frequency analysis, creating a word frequency table to discern the most pertinent features. The subsequent grouping of related words under the same feature and the selection of the most frequent term as a representative (feature keyword) ensure a focused evaluation. Language diversity in customer reviews poses another challenge. To mitigate this, a robust preprocessing step is implemented. This step addresses non-English languages, one-word comments,

spelling mistakes, and other irregularities. The goal is to transform the raw review data into structured and meaningful information, ready for sentiment analysis. Sentiment analysis is a pivotal component, breaking down customer reviews into sentences and associating sentiment scores—positive, neutral, or negative—with each. By extracting sentences that describe specific product features, the system achieves a granular understanding of customer sentiments. These sentiment scores, coupled with review votes, form the basis for the feature-level ratings. The weighting of sentiment scores and review votes is crucial for an accurate representation of customer preferences. A weighted-average methodology is introduced, acknowledging that not all opinions carry equal weight. This ensures that more significant sentiments, as reflected by review votes, have a more pronounced impact on the final feature-level ratings.

The extensive contributions of this approach are evident in its application to over 4000 mobile phones across 108 features. The novel measures, including vote-aware cumulative rating and vote-aware final rating, elevate the accuracy of sentiment score aggregation. The success is further validated through comparisons with overall customer ratings, showcasing the system's effectiveness in guiding purchasing decisions. Beyond the realm of online shopping, the proposed methodology extends to recommendation systems, emphasizing the importance of trustworthy recommendations. In the domain of movie recommendations, a hybrid recommendation system merges collaborative filtering and content-based filtering, enhanced by sentiment analysis. The goal is to cater to users' diverse tastes amidst the vast content available.

Incorporating sentiment analysis remains a central theme in the paper, acknowledging the significance of sentiment in user-generated content. The focus on Twitter as a platform underscores the real-time nature of sentiment expression, despite the inherent challenges of online data quality and the absence of ground truth. Expanding the scope to app reviews, the paper tackles the immense volume of customer feedback on digital distribution platforms. Here, the emphasis is on reducing technical debt within the technology value stream by leveraging user reviews. The proposed end-to-end recommendation framework integrates sentiment analysis, key phrase extraction, and topic recognition to identify emerging concerns and guide app developers in making informed decisions. The versatility of the proposed framework is a testament to its applicability across diverse domains. From online shopping to movie recommendations and app reviews, the common thread is the integration of sentiment analysis to extract meaningful insights. As online platforms continue to evolve, the need for sophisticated and nuanced rating systems becomes imperative. The proposed feature-level rating system, with its innovative methodologies, addresses this need by empowering consumers with detailed and informed product evaluations.

II. LITERATURE SURVEY

The landscape of online reviews has long been dominated by simplistic star ratings, often failing to capture the intricate subtleties of customer experiences. One report proposes a revolutionary shift away from these conventional metrics, introducing a pioneering approach to product evaluation. This groundbreaking system immerses itself in the realm of feature-specific sentiment analysis, delving into the wealth of customer opinions with unprecedented granularity.

At the core of this innovative approach lies a meticulous examination of customer reviews. The proposed system envisions a team of astute evaluators meticulously scrutinizing feedback, diligently identifying and dissecting mentions of specific product features, ranging from camera quality to battery life. These mentions

undergo linguistic analysis to ascertain whether the associated features are praised, criticized, or met with indifference. Notably, the inclusion of review votes, symbolized by thumbs up or down, serves as an additional layer of insight, amplifying the weight of expressed sentiments. This rigorous process culminates in the generation of individual ratings for each feature, providing a remarkably detailed and nuanced portrayal of a product's strengths and weaknesses.

The merits of this granular approach are extensive and far-reaching. Manufacturers, in particular, stand to gain unparalleled insights into the performance of their products on a feature-by-feature basis. Armed with this nuanced information, manufacturers can pinpoint specific areas for improvement, strategically prioritize development efforts, and tailor marketing strategies based on precise customer priorities. For consumers, liberated from the confines of simplified star ratings, the system opens the door to informed decision-making tailored to their specific needs. For instance, users focused on battery life can prioritize products with stellar endurance ratings, while photography enthusiasts can narrow down their choices to products with superior camera quality. Personalization takes precedence, empowering consumers to select products aligned with their unique preferences.

However, the intricate dance with customer data comes with its set of formidable challenges. The accurate identification of feature mentions and the deciphering of often-ambiguous language in reviews pose demanding tasks. Sarcasm, irony, and subtle nuances in human expression present obstacles, even for advanced algorithms. Additionally, the system's efficacy heavily relies on a wealth of review data, making it challenging to generate reliable feature ratings for products with limited feedback.

Shifting focus to the domain of entertainment, another report delves into the realm of movie recommendations with a keen eye on sentiment analysis. This research acknowledges the overwhelming amount of information on the internet and the need for reliable recommendation systems. The proposed hybrid recommendation system integrates Collaborative Filtering (CF), Content-Based Filtering (CBF), and Sentiment Analysis to deliver personalized movie recommendations. In the Collaborative Filtering component, user-movie rating data is collected, and user similarity is computed to generate recommendations based on similar users' preferences. Content-Based Filtering, on the other hand, focuses on intrinsic movie characteristics such as genres and actors, computing similarity between movies to generate recommendations with similar features. The incorporation of Sentiment Analysis, drawing on user opinions from microblogging platforms like Twitter, adds a layer of personalization by gauging user feelings towards movies.

The architecture seamlessly blends recommendations from Collaborative Filtering, Content-Based Filtering, and Sentiment Analysis using a weighted average approach. This dynamic adaptation ensures scalability and flexibility, fostering a more personalized user experience. However, the report openly acknowledges challenges, including the complexity of implementation, the data intensity involved, and the persistent issue of the cold start problem.

Despite these challenges, the proposed framework promises enhanced accuracy, scalability, and flexibility in movie recommendations. Comparative rating analysis showcases the system's ability to dynamically adjust to user preferences and integrate sentiment analysis from microblogging data, paving the way for promising avenues in personalized recommendation systems in the digital era. Shifting gears to the domain of sentiment analysis, another report provides a comprehensive exploration of the field, specifically focusing on sentiment analysis using product review data. The emergence of social media platforms has empowered users to express their opinions across various channels, including forums, micro-blogs, and social networking sites. However,

inherent flaws in online data, such as spammers and fake opinions, compromise the quality of sentiments expressed. Moreover, the absence of a consistent ground truth indicating sentiment polarity poses a significant hurdle in sentiment analysis endeavours. Addressing these challenges, the study focuses on sentiment polarity categorization, leveraging a dataset comprising product reviews from Amazon. The report proposes algorithms for identifying negation phrases, computes sentiment scores, and suggests feature vector generation methods for sentiment polarity categorization. Through experiments at both the sentence and review levels, the paper evaluates the performance of three classification models, offering a comprehensive exploration of sentiment polarity categorization in the context of product reviews.

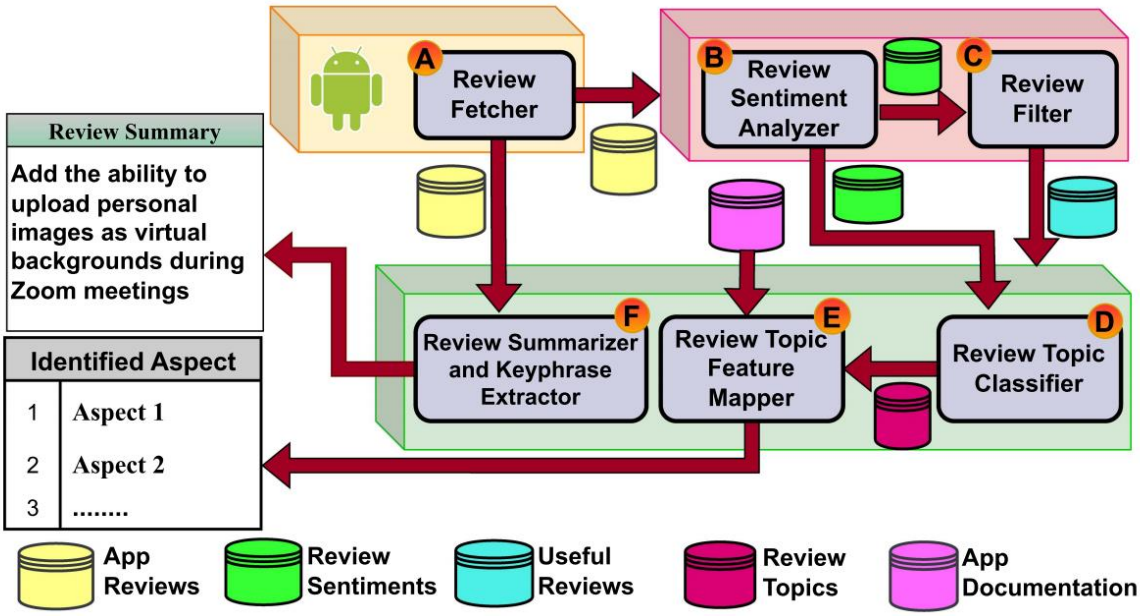


Figure 1: Proposed System for Feature-level Rating System

The proposed architecture for sentiment analysis follows a systematic and structured approach to extract insights from product reviews. The initial step involves Data Preprocessing, cleansing the data by removing irrelevant characters, symbols, and noise. The Feature Extraction stage captures relevant information, including word count, punctuation usage, sentiment words, and keywords. Model Selection becomes pivotal, considering machine learning models such as Support Vector Machines (SVM), logistic regression, and decision trees based on data characteristics and task requirements. Model Training utilizes labelled product review data to train the selected model, while Sentiment Prediction applies the trained model to new, unseen reviews to predict their sentiment. The merits of this approach are diverse, ranging from understanding customer satisfaction and identifying improvements to gaining insights into customer behaviour and enhancing product development. However, challenges include high computational costs, sensitivity to data quality and hyperparameters, and the potential for reviewer bias.

Finally, in the context of app development, another report presents a methodology for analysing app reviews to drive the technology value stream. The comprehensive research methodology begins with a thorough literature review to understand the current state-of-the-art in app review analysis, guiding the research approach based on existing insights. The report explores topic analysis techniques, including Text Classification and Key Phrase Extraction, crucial for identifying specific features discussed in user reviews. Moreover, the study emphasizes

nuanced sentiment analysis, utilizing advanced transformer models like BERT and RoBERTa to classify reviews into five sentiment classes.

The experimental evaluation phase rigorously tests the proposed framework against sixteen diverse Android applications, incorporating 3400 reviews per application from the Android Play Store. This meticulous evaluation ensures the framework's effectiveness and practical applicability. The empirical validation not only identifies areas for app improvement but also provides evidence through statistics and graphical representations, reinforcing the framework's utility in real-world scenarios. The architecture of app review analysis unfolds in a systematic process comprising pre-processing, feature extraction, analysis, prioritization, and action stages. Pre-processing involves cleaning and simplifying the raw text of customer reviews to ensure high-quality, standardized data for subsequent analysis. Feature extraction utilizes sentiment analysis tools like Stanford CoreNLP, TextBlob, and VADER, as well as topic modelling techniques like Latent Dirichlet Allocation (LDA) and keyword extraction algorithms like TextRank and RAKE, to derive meaningful insights from reviews.

The subsequent analysis phase involves a comprehensive examination of extracted features to identify improvement areas, considering sentiment distribution, common concerns, positive aspects, and critical discussion points within the reviews. Prioritization ranks improvement areas based on factors such as frequency, severity of issues, sentiment intensity, and potential user impact, directing attention towards addressing critical aspects influencing user satisfaction.

The final action phase involves the implementation of changes based on prioritized improvements, facilitated by integration with development and bug tracking tools such as Jira, GitHub, and GitLab. This integrated approach ensures prompt addressing and monitoring of identified improvements within the development lifecycle, contributing to an iterative cycle of app enhancement.

III. PROPOSED SYSTEM

The proposed system, Supervised Feature Level Rating System (SFLRS), is designed to address the challenges of today's dynamic market by providing transparent and reliable product evaluations. Leveraging advanced technologies such as machine learning, natural language processing (NLP), and data visualization, SFLRS takes a systematic approach encompassing dataset collection, splitting, preprocessing, and model training/testing. Its overarching goal is to redefine how users assess and choose products in a market flooded with options.

In this context, the system integrates Long Short-Term Memory (LSTM) networks, a type of recurrent neural network (RNN) known for their ability to analyse sequential data, particularly useful for processing textual data such as customer reviews.

Initiating with a meticulous dataset collection phase, the system aggregates product information, customer reviews, and ratings to ensure a comprehensive representation of diverse perspectives. The dataset is strategically split to facilitate granular analysis of specific product features, empowering users to make informed decisions tailored to their preferences. Subsequent data preprocessing guarantees uniformity and cleanliness, enhancing the reliability of the ensuing analysis.

A. Architecture

The proposed Supervised Feature-Level Rating System architecture is designed for a seamless user experience. The User Interface Module facilitates user interaction, while the Feature Extraction and Review Summarization

Module uses NLP and GPT tools to extract key features from reviews. The Rating Calculation Module employs a supervised learning model, addressing biases and anomalies. In this module, the LSTM model is trained and tested to predict feature-level ratings based on pre-processed data.

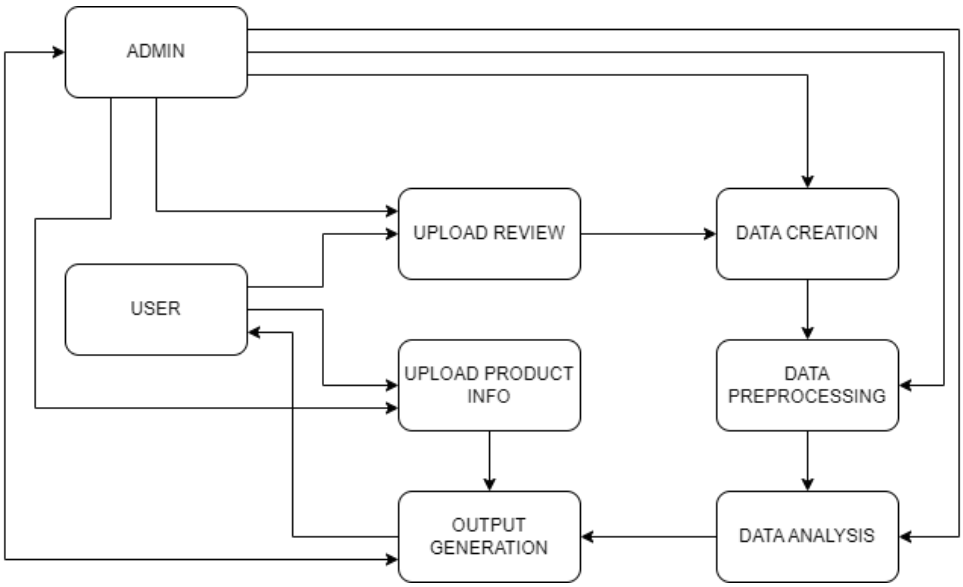


Figure 2: Architecture

The implementation of the Supervised Feature Level Rating System (SFLRS) involves integrating cutting-edge technologies such as machine learning, natural language processing (NLP), and specifically Long Short-Term Memory (LSTM) networks. This integration aims to create a robust system capable of providing transparent and reliable product evaluations in today's dynamic market landscape.

B. Dataset Collection

The first module, Dataset Collection, is a pivotal phase in building the Supervised Feature Level Rating System. This involves the meticulous gathering of product information, customer reviews, and ratings. The collected dataset serves as the foundation for subsequent analyses, capturing a broad spectrum of user experiences and perspectives. In the context of the proposal, this module aligns with the need to develop a comprehensive product rating system that relies on real-world customer feedback. By collecting diverse data, including product details and customer reviews, the system aims to provide users with a holistic view to inform their purchasing decisions. This process can involve data scraping from online platforms, API integration, or other methods to compile a rich dataset reflecting the nuances of various products.

C. Data Set Splitting

Following dataset collection, the next module, Data Set Splitting, plays a crucial role in enhancing the granularity of the analysis. Here, customer reviews are strategically divided to enable a focused examination of specific product features. This aligns with the proposed system's goal to provide users with feature-level ratings, allowing them to assess products based on their individual preferences and priorities. For instance, if a user is particularly concerned about a product's durability or design, this module ensures that the subsequent analysis can offer insights into these specific aspects. The objective is to empower users with a more nuanced understanding of products, enhancing their ability to make informed decisions.

D. Data Preprocessing

Data Preprocessing is a critical step that follows dataset collection and splitting. This module involves cleaning and normalizing the text data extracted from customer reviews. In the context of the proposed system, this step ensures that the text data is free from inconsistencies, irrelevant information, and any biases that might affect the accuracy of the subsequent analysis. By standardizing the format of the text data, the system aims to improve the overall quality and reliability of the insights generated. Techniques such as removing stop words, handling missing data, and addressing outliers may be employed to enhance the preparedness of the data for subsequent modelling.

E. Model Training & Testing

In Model Training & Testing, the LSTM model is trained on the pre-processed text data and corresponding feature-level ratings. LSTM networks excel in capturing long-range dependencies in sequential data, making them ideal for analysing textual information like customer reviews that may span multiple sentences or paragraphs. The training process involves exposing the model to historical data, enabling it to learn patterns and relationships within the dataset. During testing, the trained LSTM model predicts feature-level ratings based on new text data. The model's ability to retain context from previous words or phrases contributes to more accurate insights into product features based on customer reviews. Evaluation metrics such as Mean Squared Error (MSE) or Mean Absolute Error (MAE) are used to assess the model's performance and ensure its predictive capabilities align with the system's objectives. Integrating LSTM into the SFLRS framework enhances the system's ability to analyse textual data, providing users with more accurate and context-aware feature-level ratings derived from customer reviews.

F. Output Prediction

The final module, Output Prediction, utilizes the trained model to predict feature-level ratings. This prediction phase is the culmination of the system's efforts, providing users with insights into the strengths and weaknesses of products at a granular level. Users can benefit from these predictions to make informed decisions aligned with their preferences. The output prediction phase aligns with the proposal's emphasis on generating feature-level ratings that contribute to a comprehensive and reliable product evaluation system. It serves as a practical application of the trained model, offering users tangible and actionable information to guide their purchasing choices.

The Supervised Feature Level Rating System (SFLRS) revolutionizes product evaluations using advanced technologies like machine learning, NLP, and LSTM networks, ensuring transparent and reliable assessments in dynamic markets. SFLRS meticulously collects diverse data, preprocesses it rigorously, and employs sophisticated model training/testing, including LSTM integration, to predict accurate feature-level ratings from customer reviews and product data. This approach enables informed purchasing decisions tailored to user preferences, supported by a seamless user experience, personalized suggestions, and continuous system improvements. SFLRS's global accessibility and scalability promise mutual benefits for businesses and users, setting a new standard for transparent, reliable, and user-centric product evaluation systems.

IV. COMPARATIVE STUDY

TABLE I COMPARATIVE STUDY OF PAPERS

`

Paper	Methods/Model	*Advantages*	*Disadvantages*
"Feature-Level Rating System Using Customer Reviews and Review Votes"	GPT	Personalized Buying Decisions Improved Product	High Training Cost Feature Identification Challenge
"Movie Recommendation System Using Sentiment Analysis from Microblogging Data"	Collaborative filtering, Content based filtering	Enhanced Accuracy Scalable	Complex Implementation Data Intensive
"Driving the Technology Value Stream by Analyzing App Reviews"	BERT, RoBERTa	Comprehensive and data-driven Flexible Effective in practice	Complex and costly for some organizations Demands substantial data for NLP models
"Sentiment analysis using product review data"	REST API, Random Forest	Understand customer satisfaction and identify improvements Track sentiment trends, gain insights into customer behaviour	High computational costs Sensitivity to data quality, hyperparameters

V. CONCLUSION

Incorporating LSTM technology into the Supervised Feature Level Rating System enhances its predictive capabilities by capturing temporal dependencies within the data. LSTM models excel at analysing sequential patterns, making them invaluable in understanding how user preferences and product performance evolve over time. By integrating LSTM into the dataset collection and preprocessing phases, the system gains a deeper insight into the changing dynamics of customer reviews and product information, leading to more accurate feature-level rating predictions. This temporal understanding not only improves the system's predictive accuracy but also enhances its adaptability to shifting market trends, ensuring that users receive relevant and reliable insights for informed decision-making. The integration of LSTM technology thus strengthens the system's effectiveness in revolutionizing product evaluations and bridging the gap between consumer expectations and business offerings.

VI. REFERENCES

[1] K. R. Jerripothula, A. Rai, K. Garg and Y. S. Rautela," Feature-Level Rating System Using Customer Reviews and Review Votes," in IEEE Transactions on Computational Social Systems, vol. 7, no. 5, pp. 1210-1219, Oct. 2020, doi: 10.1109/TCSS.2020.3010807.

[2] S. Das, N. Deb, N. Chaki and A. Cortesi," Driving the Technology Value Stream by Analyzing App Reviews," in IEEE Transactions on Software Engineering, vol. 49, no. 7, pp.3753-3770, July 2023, doi: 10.1109/TSE.2023.3270708.

[3] S. Kumar, K. De and P. P. Roy," Movie Recommendation System Using Sentiment Analysis from Microblogging Data," in IEEE Transactions on Computational Social Systems, vol. 7, no. 4, pp. 915-923, Aug. 2020, doi: 10.1109/TCSS.2020.2993585.

[4] Fang, X., Zhan, J. Sentiment analysis using product review data. Journal of Big Data 2, 5 (2015).

[5] L. Yang, Y. Li, J. Wang and R. S. Sherratt," Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning," in IEEE Access, vol. 8, pp. 23522-23530, 2020, doi: 10.1109/ACCESS.2020.296985.

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

Novel Approach on Mobile Food Recognition System for Dietary Assessment

Ms.Lithiya Sara Babu*[1], Aditya Anil[2], Arjun Anil[2], Davis Sebastian[2], Sajin Sabu[2]
*[1] Assistant Professor, Department of Computer Science and Engineering ,Viswajyothi College of Engineering and Technology, Muvattupuzh, Kerala, India
[2]Department of Computer Science and Engineering ,Viswajyothi College of Engineering and Technology, Muvattupuzh, Kerala, India

ABSTRACT
The Mobile Food Recognition System is a cutting-edge app that uses deep learning to identify and analyze foods in images, providing users with immediate calorie counts and nutritional information. It also suggests healthy meal options. With its user-friendly interface, the app combines technology and nutrition science to encourage healthier eating and greater health awareness.
Keywords: CNN, DenseNet, MTL, RWL, DAN, CRN.

I. INTRODUCTION

The Mobile Food Image Recognition System for Dietary Assessment is a groundbreaking development in the realm of digital health and nutrition, heralding a new era in dietary management. This advanced mobile application is more than just a technological innovation; it's a comprehensive dietary assistant that seamlessly integrates the most sophisticated elements of deep learning and image processing technology. It's designed to revolutionize the way we approach, analyse, and understand our eating habits in a society that is increasingly focusing on health and wellness.

Central to this system's prowess is its extraordinary ability to accurately identify a diverse range of food items. This is achieved through the deployment of state-of-the-art deep learning algorithms [1], complemented by highly sophisticated image processing techniques. The application does not merely recognize the food on your plate; it delves deeper to provide a rich, detailed analysis of the nutritional profile of each item. This includes precise measurements of caloric content, protein, carbohydrates, fats, vitamins, and minerals, offering a comprehensive overview of what users are consuming. Such detailed information is crucial in empowering individuals to make more informed decisions about their diets, aligning with their health goals and nutritional needs.

Moreover, the application stands out for its innovative feature of suggesting meal options based on the foods it identifies. This unique characteristic is thoughtfully crafted to encourage not just healthy eating, but a complete

lifestyle change towards more balanced and nutritious dietary patterns. It takes into account individual preferences and dietary requirements, making it a personalized guide to healthier eating habits.

Beyond its functional capabilities, the Mobile Food Image Recognition System is a shining example of the fusion between technology and nutrition science. It's an intelligent, user-friendly tool that goes beyond passive food identification, and actively participating in users' dietary planning and management [1]. This app represents a transformative approach to dietary management, making it an essential tool for anyone looking to improve their eating habits and overall health. In essence, the application is set to become a cornerstone in the health and wellness technology sector. Its comprehensive, intuitive, and personalized approach to dietary management positions it as a vital component in the quest for a healthier, more nutritionally aware society. It's a forward-thinking solution that encapsulates the latest in technological advancements, offering a dynamic, interactive, and informative pathway to better health and nutrition.

II. LITERATURE SURVEY

A. Normalized Class Mean (NCM)

An innovative approach to personalize food image recognition by integrating the strengths of the nearest class mean (NCM) and 1-nearest neighbour (1NN) [4] classifiers with deep learning features. This hybrid model is designed to enhance adaptability to individual dietary preferences and restrictions, making it particularly suitable for applications with limited data samples. The methodology leverages deep neural networks to extract meaningful features from food images, which are then utilized by the combined classifier to improve recognition accuracy by adapting to the user-specific context of food types and preferences [4].

To evaluate the effectiveness of this approach, a unique dataset was collected from a food-logging application, reflecting the diversity and variability inherent in personal food choices. This dataset enables the demonstration of the system's superior capability in handling a wide range of food items, highlighting its potential to significantly improve personalized dietary tracking and recommendation systems. The experiments show marked improvements in accuracy over traditional methods, underscoring the benefits of personalized classifiers in environments where data scarcity and diversity pose significant challenges [4].Figure 1 shows the architecture of the Personalized Classifier System.

B. Image Segmentation & Feature Extraction

This shows a method for analysing food portions from images captured by a mobile device to estimate their nutritional value accurately [9]. Initially, users take pictures of their food from both top and side views to facilitate preprocessing and segmentation using colour and texture tools. This leads to the identification of various segments of the food portion, after which features such as size, shape, colour, and texture are extracted. These features are then classified using a Support Vector Machine (SVM) [9] to identify each food portion. The system further enhances accuracy by employing a unique calibration method using the user's thumb or an alternative object for users with thumb disabilities, which serves as a reference for measuring the real-life size of food portions.

The segmentation phase is detailed with the use of Gabor filters for texture analysis, employing a filter bank to analyse texture features around each pixel, thus enabling the system to identify various textures like soft, rough, smooth, porous, and wavy. This texture segmentation, along with colour analysis, forms the basis for accurate

food portion identification. For classification, the system utilizes an SVM with a radial basis function (RBF) [9] kernel, incorporating features extracted from the segmentation phase. The feature vectors include texture, colour, shape, and size characteristics of the food items, which are crucial for the SVM's training and classification processes.

Finally, the system calculates the volume and subsequently the mass of each food portion by applying measurements taken from the images, using the user's thumb as a reference for size calibration. This allows for the estimation of calories and nutrients by referencing nutritional tables. The system also provides options for user verification to improve accuracy, transforming it from an automatic to a semi-automatic system. This method not only offers a novel approach to dietary monitoring but also incorporates user interaction to enhance the reliability and precision of nutritional assessments.

C. Deep Learning

It is a subset of machine learning, used here to tackle the challenges associated with dietary assessment, specifically focusing on food image recognition. Deep learning enables computers to learn hierarchical representations of concepts through deep neural networks, which consist of multiple processing layers. The central problem addressed by deep learning is representation learning, where intricate concepts are expressed in terms of simpler ones. In this context, Convolutional Neural Networks (CNNs) [3], a category of deep learning models, play a pivotal role. Inspired by the visual cortex of animals, CNNs utilize learnable weights and biases in neurons, enforcing local connectivity patterns and shared parameters across the entire visual field. The three-dimensional arrangement of neurons in width, height, and depth, along with the generation of feature maps through repeated application of functions, distinguishes CNNs from traditional Multilayer Perceptrons (MLPs) [3].

The deep learning approach here, rooted in CNNs, seeks to overcome limitations associated with traditional signal processing and shallow machine learning in food image recognition [3]. Traditional methods often relied on hand-engineered features, such as SIFT and HOG, which struggled with generalizing to various food types. The recent success of deep learning, attributed to factors like large-scale annotated datasets, powerful hardware, and advanced algorithms, has opened new possibilities. We specifically focus on the application of CNNs for food image recognition, leveraging recent models like GoogLeNet and advancements like Dropout. This application of deep learning aims to enhance the accuracy and efficiency of dietary assessment by automating the recognition of food images captured by users through their mobile devices.

D. DenseNet 121

Leveraging the capabilities of a cascaded multi-head deep neural network architecture. Initially, the system employs DenseNet121 [5], a convolutional network pre-trained on the ImageNet dataset, which is then fine-tuned to adapt to the nuances of food imagery. This ensures the extraction of high-quality, discriminative features relevant to food type and state. By employing a cascading mechanism, the network effectively combines these features, enabling the precise identification of food items and their respective states.

The system further refines its recognition capabilities by integrating additional layers that specifically target the prediction of food states, such as raw, cooked, or overcooked, based on the initially identified food type. This hierarchical approach allows for a more nuanced understanding of the food items, significantly improving the accuracy and reliability of the recognition process [5]. The methodology demonstrates a notable advancement

over traditional, non-cascaded models by offering a more granular and accurate classification of food states, which holds significant potential for enhancing automated food preparation systems and providing assistance to individuals with special needs, such as the elderly or disabled, in their daily lives [5].Figure 2 shows the architecture of DenseNet 121.

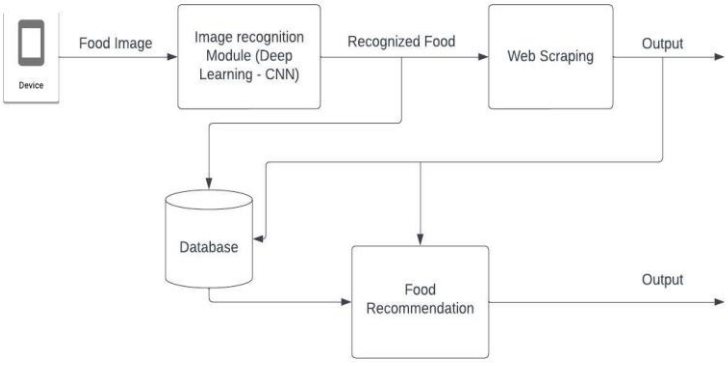
III. PROPOSED SYSTEM

The Mobile Food Image Recognition System for Dietary Assessment represents a significant technological advancement in dietary management. It combines deep learning, image processing, and nutritional analysis in an accessible format, aiming to foster healthier eating habits and a more informed understanding of nutrition and health.It uses Convolutional Neural Networks for deep learning purposes and Web Scraping techniques to obtain information from the web regarding food items and their nutritional content.

A. Architecture

Convolutional Neural Networks (CNNs) are instrumental in detecting food items and nutritional information in images, serving as a fundamental component of advanced food recognition systems. These CNNs feature intricate structures, consisting of various layers like convolutional layers, pooling layers, and fully connected layers. Each layer has a distinct function, with convolutional layers primarily responsible for detecting features in high-resolution food images. Pooling layers aid in reducing data dimensionality, and enhancing network efficiency, while fully connected layers focus on classifying and interpreting the detected features. Tailored for pattern recognition, these CNNs excel in identifying subtle visual cues in food images, such as texture variations or color discrepancies indicative of nutritional content. The hierarchical learning capacity of these networks allows them to progressively discern more complex patterns, enabling accurate and efficient recognition of food items and nutritional analysis.

Web scraping is a key technique used for database creation and update, involving specialized software or scripts that emulate a web browser's functionality to retrieve data from websites. This process is critical for extracting large volumes of information from various online sources. Once the data is collected, it undergoes a parsing process to filter and sort the relevant content, which is then systematically organized into structured formats such as CSV, Excel files, or databases. This method ensures efficient and effective gathering and updating of data, which is essential for maintaining comprehensive and current databases in various applications, including the Mobile Food Image Recognition System for Dietary Assessment. Figure 3 below shows the system architecture.



B. Implementation

In the Mobile Food Image Recognition System, Convolutional Neural Network(CNN) is used for analyzing high-resolution images of various food items, and after identifying the image web scraping is used to extract the nutrition details like calories, fats, carbs, protein vitamins minerals etc from the web. Then we can use this data to display possible healthy recipes. Also with the help of usage history, recurrently used foods are identified and a suggestion system is implemented to suggest better nutritional foods for the user to keep the diet healthy and balanced.

CNNs have been adapted to recognize patterns and textures in food images, enabling the identification of key visual cues associated with different dishes. This adaptation is crucial for accurately categorizing and analyzing the foods captured by users. The CNNs within the system excel at identifying characteristics such as discoloration, texture variations, and irregularities in food presentation. This level of detail enables the system to make precise distinctions between different dishes and ingredients, ultimately providing users with accurate information about their food intake.

Web scraping plays a vital role in attaining an up-to-date database of food-related information mainly focused on nutrition content in the food. Web scraping uses specialized software or scripts that mimic the behavior of a web browser, enabling the automated retrieval of data from a variety of online sources. These sources include nutrition databases, recipe websites, and food blogs. Once the data is collected, the web scraping process involves intricate processes for parsing and organizing the extracted information into structured formats such as CSV, Excel, or databases .

ML algorithms are used to implement a suggestion system that uses the user history, nutrition details, and the user's goal to suggest nutritional foods to take to stay healthy and achieve the goal. The app mainly provides three goals to the user.1)weight loss, 2)weight gain 3)weight maintenance.

IV. COMPARATIVE STUDY

Title	Merits	Demerits
Personalized Classifier for Food Image Recognition	1. Incremental learning 2. Personalization	1. Complexity 2. Overfitting
Measuring Calorie and Nutrition From Food Image	1.Comprehensive System Description 2. Use of Advanced Technologies	1. Complexity 2. Assumptions 3. Data availability
Food State Recognition Using Deep Learning	1. Simultaneous Prediction 3. Effective Feature Integration	1. Data Dependency 2. Overfitting Risk
A New Deep Learning-based Food Recognition System for Dietary Assessment on An Edge Computing Service Infrastructure	1. Efficient Data Processing 2. Enhanced Image Recognition	1. Reliance on Connectivity 2. Energy Consumption

V. CONCLUSION

In conclusion, the Mobile Food Recognition System For Dietary Assessment is a mobile application that uses deep learning and image processing technologies to revolutionize dietary management. By enabling accurate identification and calorie counting of various food items through user-captured images, this system significantly enhances the way individuals interact with their food. It goes beyond simple calorie tracking by providing a detailed nutritional breakdown, including proteins, carbohydrates, fats, vitamins, and minerals, thus offering a comprehensive view of dietary intake. Additionally, its ability to suggest balanced meal ideas further aids in promoting healthier eating habits. The application's user-friendly design ensures its accessibility and efficiency for a wide range of users, regardless of age or technical skill. This innovative system not only aids in making informed dietary choices but also represents a substantial leap toward fostering a more health-conscious and informed society. The Mobile Food Recognition System stands as a testament to the potential of technology in enhancing our understanding and management of nutrition and health.

VI. REFERENCES

[1] W. Min et al., "Large Scale Visual Food Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no.8,pp.9932-9949, Aug. 2023, doi: 10.1109/TPAMI.2023.3237871.

[2] J. Chen, B. Zhu, C. -W. Ngo, T. -S. Chua and Y. -G. Jiang, "A Study of Multi-Task and Region-Wise Deep Learning for Food Ingredient Recognition," in IEEE Transactions on Image Processing, vol. 30, pp. 1514-1526, 2021, doi: 10.1109/TIP.2020.3045639.

[3] C. Liu et al., "A New Deep Learning-Based Food Recognition System for Dietary Assessment on An Edge Computing Service Infrastructure," in IEEE Transactions on Services Computing, vol. 11, no. 2, pp. 249-261, 1 March-April 2018, doi: 10.1109/TSC.2017.2662008.

[4] S. Horiguchi, S. Amano, M. Ogawa and K. Aizawa, "Personalized Classifier for Food Image Recognition," in IEEE Transactions on Multimedia, vol. 20, no. 10, pp. 2836-2848, Oct. 2018, doi: 10.1109/TMM.2018.2814339.

[5] S. S. Alahmari and T. Salem, "Food State Recognition Using Deep Learning," in IEEE Access, vol. 10, pp. 130048-130057, 2022, doi: 10.1109/ACCESS.2022.3228701.

[6] B. Arslan, S. Memiş, E. B. Sönmez and O. Z. Batur, "Fine-Grained Food Classification Methods on the UEC FOOD-100 Database," in IEEE Transactions on Artificial Intelligence, vol. 3, no. 2, pp. 238-243, April 2022, doi: 10.1109/TAI.2021.3108126.

[7] H. Liang, G. Wen, Y. Hu, M. Luo, P. Yang and Y. Xu, "MVANet: Multi-Task Guided Multi-View Attention Network for Chinese Food Recognition," in IEEE Transactions on Multimedia, vol. 23, pp. 3551-3561, 2021, doi: 10.1109/TMM.2020.3028478.

[8] Z. Wang et al., "Ingredient-Guided Region Discovery and Relationship Modeling for Food Category-Ingredient Prediction," in IEEE Transactions on Image Processing, vol. 31, pp. 5214-5226, 2022, doi: 10.1109/TIP.2022.3193763.

[9] P. Pouladzadeh, S. Shirmohammadi and R. Al-Maghrabi, "Measuring Calorie and Nutrition From Food Image," in IEEE Transactions on Instrumentation and Measurement, vol. 63, no. 8, pp. 1947-1956, Aug. 2014, doi: 10.1109/TIM.2014.230353

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

Easy-G : A Complete Heart Diagnostic System

Ms.Anila Paul*[1], Farseen Muhammed[2], K A Muhammed[2], Varghese P Joseph[2]
*[1]Assistant Professor, Department of Computer Science and Engineering Viswajyothi College of Engineering and Technology, Ernakulam, Kerala, India
[2]Department of Computer Science and Engineering Viswajyothi College of Engineering and Technology, Ernakulam, Kerala, India

ABSTRACT
This conference paper affords a complete research into the detection of cardiovascular illnesses the usage of Electrocardiogram (ECG) photographs, employing ultra-modern Deep Learning and Machine Learning methodologies. Building on earlier studies, the observe targets to enhance diagnostic accuracy and performance by means of automating ECG evaluation, overcoming demanding situations related to manual interpretation. The proposed technique encompasses ECG photo preprocessing, feature extraction, and sickness category, drawing suggestion from groundbreaking studies within the area. Additionally, the paper introduces a unique fitness monitoring gadget integrating ECG, sugar stage, and blood pressure readings, prioritizing user privacy and records safety. The machine, designed for scalability and reliability, holds the capacity for integration with healthcare specialists, presenting a sensible tool for complete cardiovascular fitness tracking. Future paintings entails rigorous trying out, validation, and non-stop improvement primarily based on person feedback, aiming to bridge the distance between contemporary research and real-global health applications.

Keywords: Cardiovascular diseases, Electrocardiogram (ECG), Deep Learning, Machine Learning , Detection of Diseases

I. INTRODUCTION

Cardiovascular sicknesses (CVDs) remain a sizeable international health concern, contributing to a considerable part of morbidity and mortality global. Timely and correct detection of cardiovascular abnormalities is essential for powerful intervention and control. In recent years, advancements in clinical imaging technologies have paved the way for progressive procedures in sickness diagnosis and analysis. Among these, using Electrocardiogram (ECG) pix has won prominence for its potential to provide treasured insights into the electric pastime of the coronary heart.

This conference paper provides a comprehensive investigation into the detection of cardiovascular sicknesses via the evaluation of ECG snap shots, leveraging trendy Machine Learning (ML) and Deep Learning (DL) techniques. The research builds upon the paintings performed through Mohammed B. Abubaker and Bilal Babayi [1], as documented of their paper.

The integration of ML and DL methodologies inside the analysis of ECG pics offers a promising road for enhancing diagnostic accuracy and performance. Traditional tactics to ECG evaluation often contain manual interpretation via clinicians, which can be time-consuming and prone to human mistakes. The adoption of automated methods, powered by means of ML and DL algorithms, presents an opportunity to overcome these challenges and supply extra dependable and fast diagnoses.

This study draws suggestion from pioneering research in the discipline, such as the paintings of Smith et al. (" Automated Detection of Atrial Fibrillation Using Deep Learning with One-Dimensional Time-Series Input Data" [2]), which tested the efficacy of deep studying models in figuring out atrial traumatic inflammation patterns from ECG records.

The proposed technique encompasses pre-processing steps for ECG image enhancement, feature extraction, and the application of ML and DL fashions for disorder category. The goal is to increase a sturdy and accurate device capable of distinguishing between regular and bizarre ECG patterns, facilitating early detection of cardiovascular illnesses.

By constructing upon the foundation laid by using Abubaker and Babayi [1], this research pursuits to make contributions to the growing body of information inside the discipline of cardiovascular ailment detection. The last goal is to plan a clinically relevant tool which can useful resource healthcare specialists in making well timed and knowledgeable decisions, thereby improving affected person results and reducing the weight of cardiovascular diseases on public health

II. LITERATURE SURVEY

A. Deep Learning Based Cardiac Navigation

The study described in reference [3] concentrates on enhancing the clinical evaluation of cardiac vitality by employing time-continuous tomographic imaging of the heart. Particularly using magnetic resonance imaging (MRI) with multiple imaging contrasts. The conventional approach involves external sensors like electrocardiograms (ECG) for navigation, introducing additional workflow complexities. The paper introduces a navigation system devoid of sensors, relying instead on deep learning method that eliminates the need for manual feature extraction. A classifier is trained to directly estimate R-wave time points from the imaging data, enabling continuous cardiac MRI without relying on ECG signals. The method is assessed using 3-D protocols for in-vivo continuous cardiac MRI under free-breathing conditions, encompassing both single and multiple imaging contrasts. The findings reveal an accuracy exceeding 98% on subjects not encountered before, and the image quality is comparable to the state-of-the-art ECG-based reconstruction. This innovative approach provides an ECG-free workflow for continuous cardiac scans, enabling simultaneous anatomic and functional imaging with multiple contrasts. Additionally, it holds the potential for seamless integration into various continuous MRI sequences without necessitating modifications to the sampling scheme.

B. Deep Neural Network Based On ECG

The paper [4] addresses the challenges in deciphering large datasets for cardiovascular disease (CVD) diagnosis and treatment using electrocardiograph (ECG) signals. While ECG is a valuable tool, processing vast datasets remains a hurdle. The study introduces a strategy employing a deep neural network (DNN) that undergoes a learning stage to enhance classification accuracy through robust feature extraction. Following this, a genetic

algorithm (GA) process is employed to identify the optimal combination of feature extraction and classification. Comparative analysis with state-of-the-art methods highlights significant improvements, as the proposed technique shows a notable increase in average accuracy and F1 score. These results indicate the model's potential as an analytical module, capable of alerting users or medical experts when anomalies are detected. This contributes to more effective and accurate diagnoses in the context of Cardiovascular Diseases (CVD).

C. A Naive Bayes Classifier

In reference [5], a fully integrated electrocardiogram (ECG) signal processor (ESP) is presented, specifically designed for predicting ventricular arrhythmia. This ESP incorporates a distinctive set of ECG features and employs a naive Bayes classifier for real-time and adaptive detection and delineation of the PQRS-T waves. This approach ensures robustness to variations in the ECG signal, achieving high sensitivity and precision. To validate its effectiveness, the study utilizes two databases from MIT Physio Net and the American Heart Association, demonstrating an overall classification accuracy of 86% on out-of-sample validation data with a 3-second window size.

The proposed ESP architecture is implemented using a 65-nm CMOS process, occupying a compact 0.112-mm² area, and consuming 2.78-μW power at an operating frequency of 10 kHz and an operating voltage of 1 V. Notably, this ESP represents the first ASIC implementation for ECG-based prediction of ventricular arrhythmia up to 3 hours before onset, showcasing its potential as an advanced and efficient tool for early arrhythmia detection.

D. Feature Extraction

The focus of the study detailed in reference [6] is on the timely identification of Hypertrophic Cardiomyopathy (HCM), a cardiovascular disease, utilizing electrocardiograms (ECG). Introducing a dedicated A classifier specifically designed for cardiovascular patients with the specific goal of identifying HCM patients from standard 10-second, 12-lead ECG signals, the classifier distinguishes individual heartbeats characteristic of HCM, using non-HCM heartbeats as controls. Extracting a comprehensive set of 504 morphological and temporal features from ECG signals for heartbeat classification, encompassing a combination of features that includes both widely adopted and recently developed elements, the investigation utilizes classifiers such as random forest and support vector machine. Through 5-fold cross-validation, these classifiers surpass a logistic regression classifier, achieving precision around 0.85, with recall (sensitivity) and specificity at approximately 0.90. Feature selection experiments reveal that the performance of the classifier can be matched by a subset of 264 highly informative features, equivalent to using the complete set. These findings underscore the potential of the proposed classifier for precise identification of HCM patients and emphasize its efficiency in feature selection for streamlined analysis.

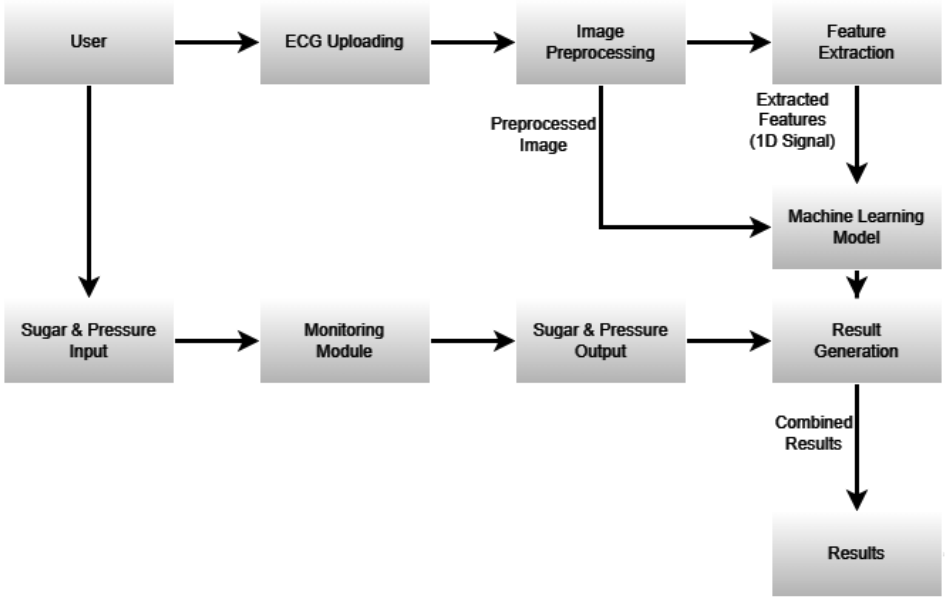
III. PROPOSED SYSTEM

The proposed health monitoring system integrates ECG, sugar level, and blood pressure readings for a comprehensive health assessment. Users input their readings through a user-friendly interface, initiating data flow to a modified CNN model. This model incorporates separate branches for ECG, sugar level, and blood pressure data, leveraging predefined thresholds for categorization. The combined results from ECG analysis and

individual parameter assessments yield an overall health status, presented through a clear and visually intuitive interface. The system includes interpretability features to explain the model's assessments and provides alerts and recommendations for high or low parameter values. Emphasis is placed on privacy, security, and compliance with health data regulations. User feedback and regular model updates contribute to continuous improvement. The deployment is scalable and reliable, with the potential for integration with healthcare professionals for further analysis and consultation, while comprehensive documentation ensures usability and understanding for both users and developers.

A. Architecture

The proposed health monitoring system architecture comprises three main components: ECG module, Sugar Level module, and Blood Pressure module, integrated into a unified system. The ECG module utilizes a lightweight CNN model specifically designed for analysing 12-lead ECG images. The Sugar Level and Blood Pressure modules rely on predefined thresholds and a simple decision-making process to categorize readings as high, low, or normal. These modules collectively feed their results into a central module, which integrates the outputs, interprets the combined health status, and presents it through a user-friendly interface. The system prioritizes user privacy and data security, implementing encryption and secure data transmission protocols. A feedback loop allows users to provide additional information or context, facilitating continuous improvement of the model. The architecture is designed for scalability, accommodating potential future additions and enhancements. Regular updates and maintenance ensure the system's reliability and effectiveness in providing users with a holistic health assessment. Integration with healthcare professionals and adherence to regulatory standards enhance the overall functionality and trustworthiness of the system.



B. Implementation

- **Data Collection**: Gather a diverse dataset containing ECG images, blood pressure readings, and sugar level readings. This dataset should cover a range of health conditions to train and test the system effectively.
- **Preprocessing:**
  - **ECG Module:** Preprocess ECG images by cropping, resizing, and augmenting to ensure uniformity. Implement the lightweight CNN model for ECG classification.

- **Sugar Level and Blood Pressure Modules:** Define threshold values for categorizing sugar and blood pressure levels as high, low, or normal.

- **System Integration:**
  - Develop a central module to integrate the ECG, sugar level, and blood pressure modules.
  - To guarantee the privacy and security of data, it is essential to implement secure data transmission protocols for user health data.

- **Decision-Making Logic:**
  - Design decision-making logic that interprets outputs from individual modules to provide an overall health status.
  - Consider feedback mechanisms to allow users to provide additional context or correct misinterpretations.

- **User Interface:**
  - Develop a user-friendly interface to display health status, including ECG results, sugar level, and blood pressure readings.
  - Ensure the interface is accessible and easy to understand for users of varying technical backgrounds.

- **Scalability and Maintenance:**
  - Design the system to be scalable, allowing for the addition of new features or modules in the future.
  - Implement regular updates and maintenance routines to address bugs, improve performance, and enhance the system's capabilities.

- **Security Measures:**
  - Incorporate encryption and authentication mechanisms to safeguard user data.
  - Comply with relevant data protection and privacy regulations.

- **Testing and Validation:**
  - Conduct rigorous testing to validate the accuracy and reliability of the health monitoring system.
  - Test the system with diverse datasets and real-world scenarios to ensure robustness.

- **User Training:**
  - Provide user training materials to help individuals understand how to use and interpret the information provided by the system.

- **Deployment:**
  - Deploy the system in a controlled environment initially, gather user feedback, and make necessary adjustments.
  - Gradually expand the deployment to a wider user base.

IV. COMPARATIVE STUDY

References	Title	Techniques	Merits	Demerits
[3]	Deep learning-based ECG free Cardiac Navigation for Multi-Dimensional and Motion-Resolved Continuous Magnetic Resonance Imaging.	• Deep learning based cardiac navigation • MRI	• ECG-free • Imaging that encompasses multiple dimensions and involves motion-resolved techniques. • Fully automatic Potentially applicable to other continuous sequences	• Requires deep learning model training • Performance may depend on the training data • Limited clinical experience
[4]	A Multi-tier Deep Learning Model for Arrhythmia Detection	• Deep Neural Network based on ECG • Genetic Algorithm	• Promising Results: • Robustness: • Multi-Tier Deep Learning Model:	• Limited Information: • Lack of Detailed Analysis of Comparison:
[5]	Low-Power ECG Based Processor for Predicting Ventricular Arrhythmia	• A naive Bayes classifier • Identification and delineation of the PQRS-T waves.	• Energy Efficient: • Integrated Design: • High Accuracy: • Simple Design: • Compact Size: • Superior Performance:	• Challenging to implement or calculate. • Less specificity • Late detection interval
[6]	Utilizing ECG based Heartbeat Classification for Hypertrophic Cardiomyopathy Identification	• Feature extraction • Feature selection • ML classifiers	• High performance • Feature Selection • Information Gain Criterion	• Discretization of Continuous Features • Limited Dataset • Baseline Classifier Performance

V. CONCLUSION

In conclusion, this conference paper presents a concise overview of relevant literature in cardiovascular health monitoring and introduces an innovative health assessment system. Inspired by advancements such as ECG-free

continuous cardiac MRI [2], a neural network for large dataset diagnosis [3], an ECG signal processor for arrhythmia prediction [4], and an HCM classifier [5], our proposed system integrates ECG, sugar level, and blood pressure readings. The architecture prioritizes user privacy, data security, and compliance with regulations, offering interpretability features and alerts. The scalable and reliable deployment is designed for user engagement and potential integration with healthcare professionals. Future work involves rigorous testing, validation, and continuous improvement based on user feedback, with the ultimate goal of providing a practical tool for comprehensive cardiovascular health monitoring. This paper serves as a bridge between cutting-edge research and real-world health applications.

VI. REFERENCES

[1] M. B. Abubaker and B. Babayigit, ”Detection of Cardiovascular Diseases in ECG Images Using Machine Learning and Deep Learning Methods,” in IEEE Transactions on Artificial Intelligence, vol. 4, no. 2, pp. 373-382, April 2023, doi: 10.1109/TAI.2022.3159505.

[2] J. Smith et al., "Automated Detection of Atrial Fibrillation Using Deep Learning with One-Dimensional Time-Series Input Data," IEEE Transactions on Biomedical Engineering, 2019.

[3] E. Hoppe et al., ”Deep Learning-Based ECG-Free Cardiac Navigation for Multi-Dimensional and Motion-Resolved Continuous Magnetic Resonance Imaging,” in IEEE Transactions on Medical Imaging, vol. 40, no. 8, pp. 2105-2117, Aug. 2021, doi: 10.1109/TMI.2021.3073091.

[4] M. Hammad, A. M. Iliyasu, A. Subasi, E. S. L. Ho and A. A. A. El Latif, ”A Multitier Deep Learning Model for Arrhythmia Detection,” in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-9, 2021, Art no. 2502809, doi: 10.1109/TIM.2020.3033072.

[5] Low-Power ECG-Based Processor for Predicting Ventricular Arrhythmia [Nourhan Bayasi, Member, IEEE, Temesghen Tekeste, Member, IEEE, Hani Saleh, Senior Member, IEEE, Baker Mohammad, Senior Member, IEEE, Ahsan Khandoker, Senior Member, IEEE, and Mohammed Ismail, Fellow, IEEE et.al, 2016]

[6] Utilizing ECG based Heartbeat Classification for Hypertrophic Cardiomyopathy Identification [Quazi Abidur Rahman, Larisa G. Tereshchenko, Matthew Kongkatong, Theodore Abraham, M. Roselle Abraham, and Hagit Shatkay et.al, 2015]

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

Elevating Crisis Response: Machine Learning and Cutting-Edge Technologies for Disaster and Pandemic Management

Ms. Arsha J K[1], Diya Xavier[2], Emlin Maria Roy[2], Grace John[2], Sona Sunny[2]
*[1]Assistant Professor, Department of Computer Science, VJCET, Ernakulam, Kerala, India
[2]Student, Department of Computer Science, VJCET, Ernakulam, Kerala, India

ABSTRACT
The *Resource Recovery Planner* project aims to harness cutting-edge machine learning algorithms for effective disaster and pandemic management. Acknowledging the importance of technologies like IoT, object sensing, UAVs, 5G, cellular networks, smartphone-based systems, and satellite-based systems, the project explores their integration to address these challenges. By emphasizing machine learning's adaptability to handle extensive, multi-dimensional datasets, the project investigates applications in disaster management, including predicting occurrences, optimizing crowd evacuation routes, and analyzing real-time social media data. Additionally, it underscores the relevance of machine learning in pandemic scenarios, covering predictive modeling, spread monitoring, and disease diagnosis. The project involves training machine learning models using historical data to enhance prediction accuracy and explores synergies with other technologies for robust disaster and pandemic management. While acknowledging challenges related to data quality and ethical considerations, the project outlines future research directions.

Key Words: Support Vector Machine, Random Forest Algorithm, Convolutional Neural Network.

I. INTRODUCTION

The escalating global population growth and impacts of climate change are intensifying the occurrence and severity of natural disasters is posing significant challenges to communities worldwide. These events result in extensive infrastructure damage, loss of life, and disruptions to economies. Against this backdrop of mounting environmental threats, there is an urgent need for effective disaster preparedness and response systems. So, our project aims to create a Resource Recovery Planner that uniquely integrates real-time weather data with historical records of past disaster occurrences. Integrating advanced machine learning algorithms, the system predicts potential dangers. By harnessing live weather updates from community volunteers and insights from historical disaster patterns, the planner facilitates proactive measures such as establishing evacuation routes, ensuring essential resource availability, and coordinating emergency services. The proposed Resource Recovery Planner operates at the intersection of technology and disaster management, ushering in a transformative approach to community preparedness and response.By harnessing machine learning, it analyzes extensive datasets, providing invaluable predictive capabilities amidst uncertainties. This foresight enables informed

decision-making and empowers communities to take preventive measures, mitigating disaster impacts. Rooted in real-time data and historical insights, the planner embodies a holistic approach, equipping communities with adaptable tools to navigate complex natural challenges and foster resilience in our ever-changing world. This process represents a pivotal initiative in the pursuit of enhanced disaster resilience, amalgamating cutting-edge technologies and data-driven methodologies. By addressing the urgency of the situation through the creation of a resource recovery planner, it aspires to redefine how communities prepare for and respond to natural disasters, ultimately contributing to a more resilient and adaptive global response to the complex interplay of environmental and demographic factor

II. LITERATURE SURVEY

A. ConvLSTM HYBRID ALGORITHM

The research[1] introduces the Flood Index (IF) as a key tool for assessing flood conditions, representing a standardized form of Effective Precipitation (PE). IF is designed to account for the cumulative impact of both current and previous day's precipitation, considering the gradual diminishment of the influence of past rainfall by hydrological factors. While IF itself cannot predict future flood events, the study aims to address this limitation by developing a predictive model for the Flood Index.

The research proposes a hybrid deep learning approach for flood forecasting, combining Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) Network. This fusion aims to leverage the strengths of both networks, striving for an efficient, robust, and accurate predictive model. The emphasis extends beyond model creation to evaluating its predictive capabilities across different forecast horizons.

To validate the proposed ConvLSTM model, the research conducts rigorous testing in flood-prone regions of Fiji, a location frequently affected by floods. The model undergoes testing against nine distinct rainfall datasets, providing a comprehensive assessment of its performance. The inclusion of historical and real-time rainfall data enables the model to capture the evolving Flood Index over time, enhancing its forecasting accuracy.

In summary, the research introduces IF as a pivotal tool for flood assessment, acknowledges its limitation in predicting future events, and addresses this by proposing a hybrid deep learning model (ConvLSTM) for flood forecasting. The validation process involves testing the model in flood-prone regions of Fiji using multiple rainfall datasets to assess its performance and forecasting accuracy across different forecast horizons.

This study introduces an innovative hybrid deep learning model for flood prediction, validates its effectiveness in real-world scenarios, and underscores its potential to enhance flood risk management and adaptation strategies.

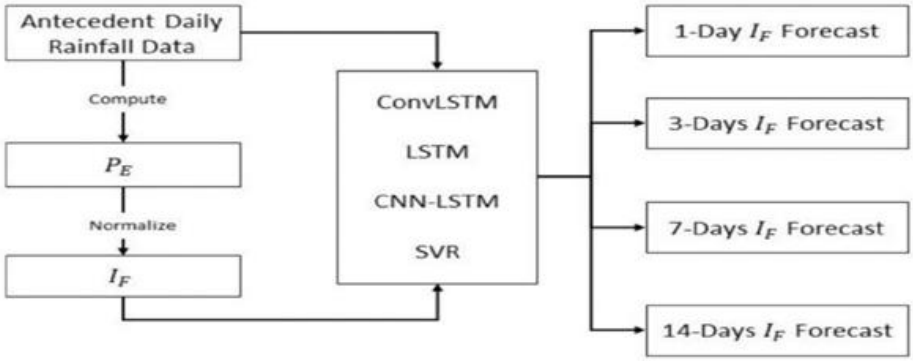


Fig. 1.proposed system LS

B. FFM using FEDERATED LEARNING

The Flood Forecasting Model (FFM), as proposed in [2], stands as an advanced system tailored for predicting floods across diverse rivers and barrages within a specified region. Its intricacy lies in its five-layer architecture, where each layer plays a distinct role in the flood prediction process. Starting with the Physical Layer situated at the system's edge, this layer serves as the initial data collection point. Outfitted with sensors, it captures multidimensional data encompassing water levels, weather conditions, and other pertinent parameters. Subsequently, this information is transmitted to local client stations.

The next stage is the Preprocessing Stage, which is crucial before local model training. This phase, occurring at the client stations, ensures data accuracy and reliability. Key preprocessing steps involve dataset normalization, removal of corrupt records, and addressing missing values. Through meticulous data preparation, the FFM lays the groundwork for robust flood predictions, contributing significantly to effective flood risk management and the formulation of adaptation strategies.

Moving to the network layer, the model focuses on the secure transfer of local data models from the client stations to the central server. This layer plays a crucial role in facilitating efficient communication between the edge devices and the central processing unit. The subsequent processing layer completes the Federated Learning (FL) cycle, embodying a holistic approach to flood prediction. Local models are trained at client stations, transmitted to the central server, aggregated, and utilized to train a global data model. The central server, acting as a hub for data collaboration, employs innovative techniques such as the Federated Averaging Algorithm (FedFlood) during model training. This approach, involving eighteen clients, enhances the model's predictive capabilities by leveraging diverse local datasets, thereby creating a more comprehensive and accurate flood prediction system.

Apart from the structural complexities of the Flood Forecasting Model, the article highlights essential parameters relevant to the application of Federated Learning in this context. These parameters include the batch size of the local model (denoted as b), the learning rate for the local dataset, the number of training rounds (e), threshold criteria, and the time required for each training round (t). These factors play crucial roles in achieving an effective Federated Learning process.

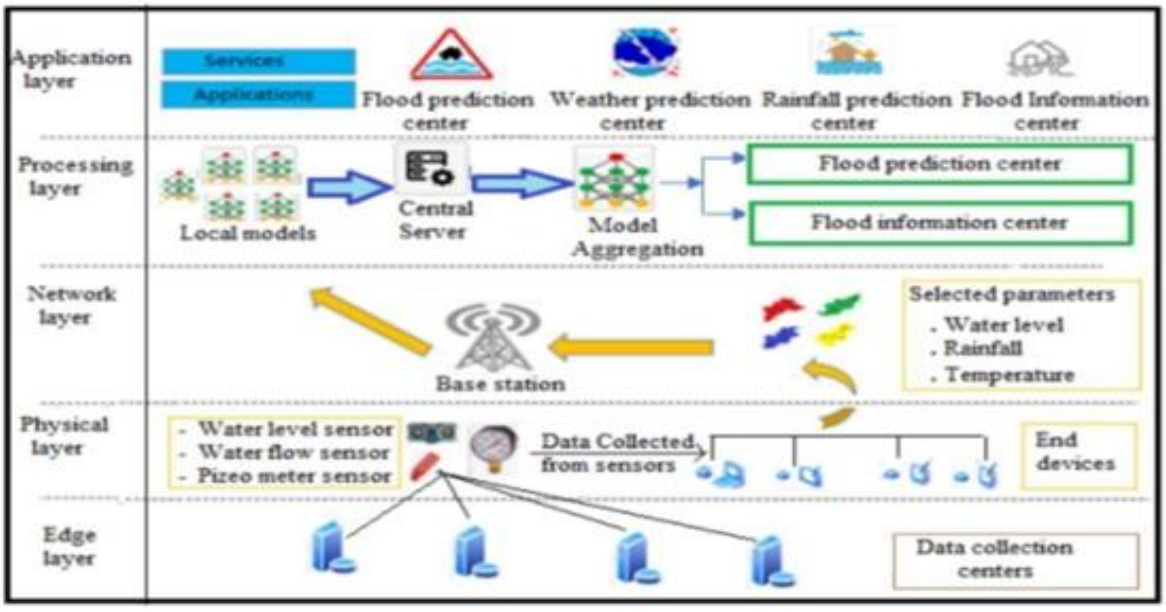


Fig. 2.proposed system FFM

What sets this proposed model apart is the integration of two-dimensional Federated Learning. In the first dimension, local data models are trained on individual devices. Subsequently, a global data model is trained based on the aggregated information from these local models in the second dimension. This nuanced approach enhances the model's efficiency and accuracy in forecasting floods within the specified region.

C. Big Data and Emergency Management: concepts

Big Data and Emergency Management (BDEM)[3] involves advanced techniques to collect, analyze, and interpret large and intricate datasets. The primary goal of BDEM is to enhance emergency response, recovery, and mitigation. It achieves this by screening critical and diverse information, predicting relief demands, aiding decision-makers, and constructing real- time knowledge systems. Mathematical models play a crucial role in optimizing relief decision-making, including evacuation scheduling, resource allocation, and logistics. However, challenges persist, such as dealing with the complexity of modern cities, predicting human behavior, and collecting long-term sensing data.

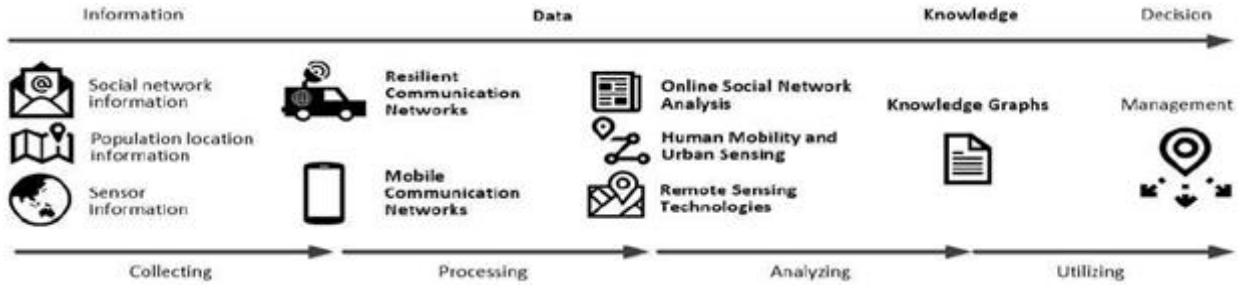


Fig. 3.overview BDEM system

Methods and Technologies

1. Remote sensing technologies: These include satellite imagery, aerial photography, and other forms of remote sensing to collect data on disaster-affected areas.
2. Resilient communication networks: These are communication networks that are designed to withstand disasters and continue to function during emergencies.
3. Mobile communication networks: These are communication networks that are designed to be mobile and flexible, allowing emergency responders to communicate and coordinate in real time.
4. Human mobility modeling: This involves using big data to model and predict human mobility patterns during emergencies, which can help emergency responders allocate resources more effectively.
5. Urban sensing: This involves using sensors and other data collection methods to monitor urban environments and detect changes that may indicate an emergency.
6. Online social network analysis: This involves analyzing social media data to gain insights into public sentiment, identify potential risks, and monitor relief efforts.
7. graphs: These are graphical representations of knowledge that can be used to identify relationships between different data points and gain insights into complex systems.
8. Edge and fog computing: These are distributed computing paradigms that allow data to be processed and analyzed closer to the source, reducing the need for data transmission and improving response times.

D. Flood Prediction Integrating Volunteer Data

Floods cause extensive damage to life and property. As a response to this, the paper proposes a machine-learning framework by combining weather, water flow, geographic and volunteer data to generate disaster response and preparedness strategies. Flood-predictive elements like duration of precipitation, river networks, altitude relative to sea level, slope of terrain, land utilization and coverage, recurrent flooding, and water flow orientation could be used to anticipate the occurrence of flood in advance. The data acquisition for flood forecasting consisted of several stages :-Weather data forecast, collection of geographic data sourced from Global Flood Awareness System, TMD Big Data platform was used to get the hourly prediction of rain and Real-time data generation.

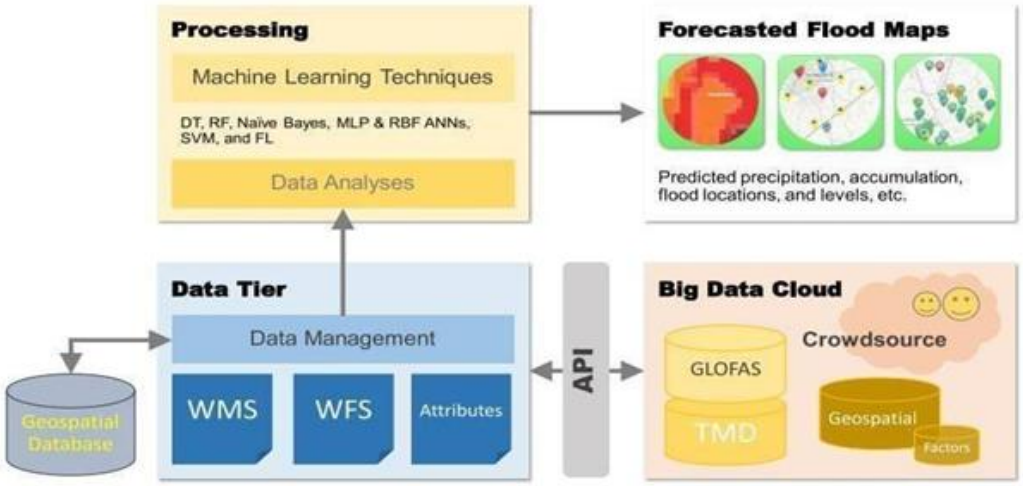


Fig. 4.overview

The collected data were stored in a geodatabase and subsequently analyzed using one of the contemporary machine learning approaches. Data exchanges were facilitated through interface technologies like Web Map Service (WMS), Web Feature Service (WFS), TMD API, and Google Maps API.Various cross-platform APIs were used to gather big data from prominent agencies such as TMD. The resultant model incorporating both the ML strategies and the four data categories could help in the detection of floods. The ML frameworks that were used to choose the best prediction model were RF, ANN ,SVM, DT (J48),Fuzzy Logic and Na¨ıve Bayes. MLP ANN, SVM, and RF had higher classification accuracies of 97.83%, 96.67%, and 96.67% respectively while determining the performance of the model.The resultant model regularly updates its prediction which contributed to improving its accuracy.

III. PROPOSED SYSTEM

A. Architecture

The design of the flood prediction and management system is meticulous, leveraging both real-time live data and historical rainfall data. Through the integration of these datasets, the system forms a comprehensive training dataset for machine learning models. In particular, linear regression models are utilized to discern patterns and relationships in the data. This machine learning framework empowers the system to forecast flood occurrences by analyzing both current and historical weather conditions. In response to an elevated risk of

flooding indicated by the model, the system dynamically updates critical information, such as evacuation routes and shelter locations.

The strength of this integrated approach lies in the system's ability to adapt to changing conditions. By combining real- time inputs from volunteers with historical trends, the model becomes skilled at recognizing subtle patterns and continuously improving its accuracy. The active involvement of volunteers in data collection further enhances the system's effectiveness, ensuring that the dataset remains up-to-date and reflective of current environmental conditions. The dynamic updates to evacuation routes and shelter locations are crucial for managing flood situations, allowing authorities and the public to make informed decisions promptly. This real-time responsiveness contributes to effective disaster preparedness and response. Overall, the flood prediction and management system not only relies on machine learning for accurate predictions but also emphasizes the collaborative efforts of volunteers, resulting in a robust and valuable tool for proactive flood prediction and adaptive management strategies.

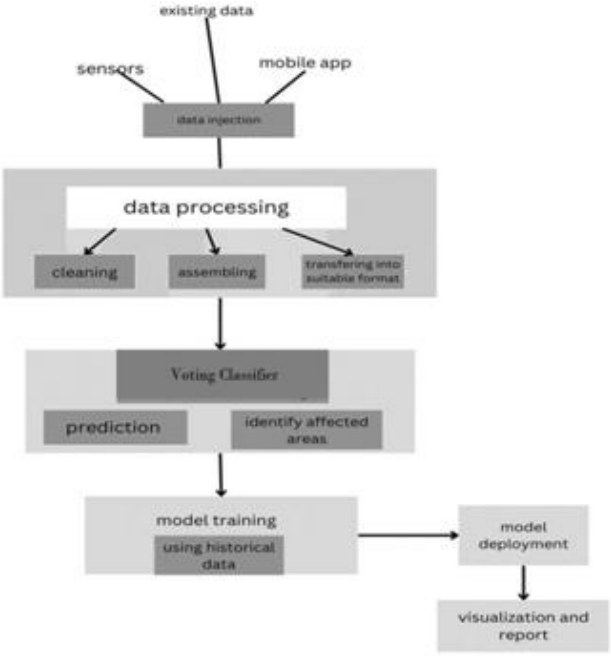


Fig. 5. Architecture diagram

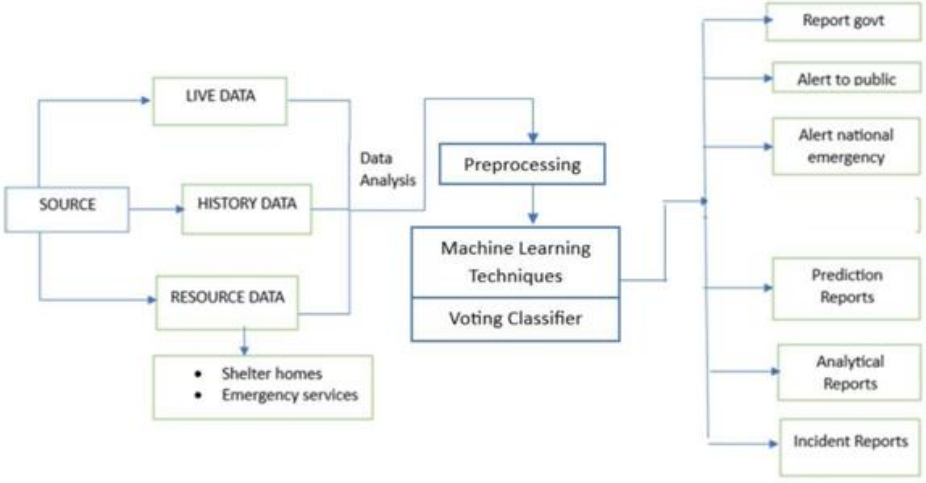


Fig 6. Proposed system overview

B. Implementation

Data Ingestion: This involves gathering information from various sources, such as sensors, social media, and satellite imaging. The collected data is then stored in a central database or data lake.

Data Pre-processing: Raw data needs preparation before analysis. This stage includes tasks like data cleaning, formatting, and feature engineering. Normalization and handling missing values are also part of this process.

Machine Learning Models: These models serve diverse purposes, such as identifying affected areas, forecasting disaster occurrences, and assessing damage levels. Techniques like clustering, regression, and classification are employed during model development.

Model Training: Historical data is used to train machine learning models. This step often requires substantial computing resources. Distributed training methods and coordination with tools like Kubernetes, along with containerization using platforms like Docker, play a role in this phase.

Integration with Decision Support Systems: Disaster management authorities incorporate machine learning forecasts into decision support systems. This ensures that the insights generated by machine learning are effectively utilized in decision- making processes.

Feedback Loop: Establishing a feedback loop over time to enhance machine learning model performance. This involves monitoring model effectiveness, collecting fresh data, regularly retraining models, and releasing updated versions.

Visualization and Reporting: Decision-makers can gain a clearer understanding of the current situation and organize recovery efforts by utilizing dashboards, maps, charts, and other visual representations of machine learning analysis results.

IV. COMPARATIVE STUDY

TABLE I. COMPARATIVE STUDY OF LITERATURE SURVEY

Model	Accuracy	Advantages	Disadvantages
ConvLSTM Hybrid Algorithm [1]	94%	Flexibility	Model Complexity
FFM using Federated Learning [2]	93.4%	Data privacy	Data synchronization is complex
Big Data and Emergency Management [3]	99.5%	High Accuracy	Data Quality and Labelling Issues
Flood Forecasting System Based on Integrated Big and Crowdsource Data [4]	90%	Improves the effectiveness of disaster relief efforts	Difficulty in maintaining

V. CONCLUSION

In conclusion, effective disaster management necessitates thorough preparedness and strategic planning, with a central focus on creating and executing emergency response strategies. The bedrock of a resilient disaster management framework lies in establishing robust early warning systems. These systems play a pivotal role by providing timely and accurate information to communities facing imminent threats, facilitating well-organized evacuations of individuals and assets. Proactive implementation of such measures can significantly mitigate potential loss of life and property. Additionally, clear communication channels and collaborative efforts among stakeholders are essential for enhancing overall response and recovery endeavors. These communication pathways act as vital conduits for disseminating critical information and seamlessly coordinating efforts. The efficiency of these communication mechanisms ensures effective mobilization of resources and well-coordinated response efforts.

VI. REFERENCES

[1] M. Moishin, R. C. Deo, R. Prasad, N. Raj and S. Abdulla, ”Designing Deep-Based Learning Flood Forecast Model With ConvLSTM Hybrid Algorithm,” in IEEE Access, vol. 9, pp. 50982-50993, 2021, doi: 10.1109/ACCESS.2021.3065939.

[2] M. S. Farooq et al., ”FFM: Flood Forecasting Model Using Federated Learning,” in IEEE Access, vol. 11, pp. 24472-24483, 2023, doi: 10.1109/AC- CESS.2023.3252896.

[3] X. Song et al., ”Big Data and Emergency Management: Concepts, Methodologies, and Applications,” in IEEE Transactions on Big Data, vol. 8, no. 2, pp. 397-419, 1 April 2022, doi: 10.1109/TBDATA.2020.2972871.

[4] T. Schempp, M. Hong, H. Zhang, R. Akerkar and A. Schmidt, ”An Integrated Crowdsourced Framework for Disaster Relief Distribution,” 2018 5th International Conference on Information and Communication Technologies for Disaster Management (ICT-DM), Sendai, Japan, 2018, pp. 1-4, doi: 10.1109/ICT-DM.2018.8636372.

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

Intelligent Traffic Control in Multi-Junction Scenarios: A YOLOv8 - based Approach

Ms. Bency Cleetus*[1], Gayatri P G[2], Geethukrishna T S[2], Sajitha Francis[2], Serene John[2]
*[1]Assistant Professor, Department of Computer Science, VJCET, Ernakulam, Kerala, India
[2]Student, Department of Computer Science, VJCET, Ernakulam, Kerala, India

ABSTRACT
Video-Based Vehicle detection and Counting for traffic control aims to address traffic congestion and enhance the overall efficiency of urban intersections. By dynamically adjusting traffic signals based on real-time traffic conditions, our approach seeks to improve traffic flow and minimize congestion. The integration of YOLOv8 ensures reliable and efficient object detection, enabling the system to effectively monitor and analyze various vehicle types. The framework introduces a novel system for traffic control at urban junctions through the implementation of a Video-Based Vehicle Detection and Counting mechanism. Leveraging the capabilities of YOLOv8 for real-time object detection, our system strategically deploys high-resolution cameras to accurately identify and classify diverse types of vehicles traversing through junctions. The collected data, comprising vehicle counts and types, is centrally processed, and a dynamic traffic control algorithm utilizes this information to optimize signal timings in real-time.This outlines the architecture of our Video-Based Vehicle Detection and Counting system, emphasizing the use of YOLOv8 for robust real time object detection. The system is able to accurately count and classify vehicles and provide valuable data for the dynamic traffic control algorithm. The proposed solution presents a promising avenue for optimizing urban traffic management, contributing to reduced congestion and improved overall traffic efficiency at intersections.

Index Terms: YOLOv8, Vehicle Detection, Object Detection, Vehicle counting.

I. INTRODUCTION

In the dynamic realm of transportation systems, the imperative need for precise vehicle counting has escalated, serving as a foundational element for the implementation of intelligent traffic solutions. The accuracy of traffic flow information, encompassing parameters like lane occupancy, congestion levels, and real-time traffic status, underpins the formulation of effective traffic control strategies, timely accident warnings, and the optimization of dynamic signal timings. Traditional methodologies, reliant on sensors such as magnetic coils or microwave detectors, are constrained by limitations in acquiring detailed information and burdened with high installation costs. In response to these challenges, video-based vehicle counting systems have emerged as a compelling solution, harnessing the advancements in image processing technology.

The paradigm of machine vision-based vehicle counting unfolds through a tripartite process, involving detection, tracking, and trajectory processing. The initial phase, object detection, aims to precisely locate and classify vehicles within images. While earlier approaches relied on manually engineered features like SIFT, HOG, and Haar-like, coupled with classifiers such as SVM and Adaboost, recent strides in deep learning, particularly Deep Convolutional Neural Networks (DCNN), have ushered in a new era of heightened accuracy. This paper strategically capitalizes on this progress, leveraging YOLOv8 for robust and efficient object detection. The proposed project envisions the implementation of an intelligent traffic signal control system that dynamically adjusts signal timings based on live vehicle count data. Such a system possesses the capability to adapt signal durations, prioritizing the side of the intersection with higher vehicle counts. By deploying this system, cities can achieve a more responsive and adaptive traffic signal control mechanism, effectively managing congestion, reducing delays, and enhancing overall traffic efficiency. The intelligent allocation of signals based on real-time vehicle counts contributes to the development of a dynamic and optimized traffic control strategy, marking a significant advancement in intelligent transportation systems.

II. RELATED WORKS

Traffic control research is vital for improving urban mobility by developing intelligent systems that reduce congestion and enhance transportation efficiency. It plays a key role in addressing the challenges of growing urbanization and the need for sustainable, effective traffic management solutions.

A. YOLO V3 and KCF-Based Vehicle Counting in Intelligent Transportation

The research paper [1] addresses the critical task of estimating the number of vehicles in a traffic video sequence, a key element in Intelligent Transportation Systems (ITS). Traditional methods for vehicle counting rely on specialized sensors like magnetic coils, microwave, or ultrasonic detectors, which are limited by factors such as cost, installation complexity, and the information they provide. With advancements in image processing technology, video-based vehicle counting systems have emerged as a more cost-effective solution, offering additional parameters such as vehicle category, density, and speed.

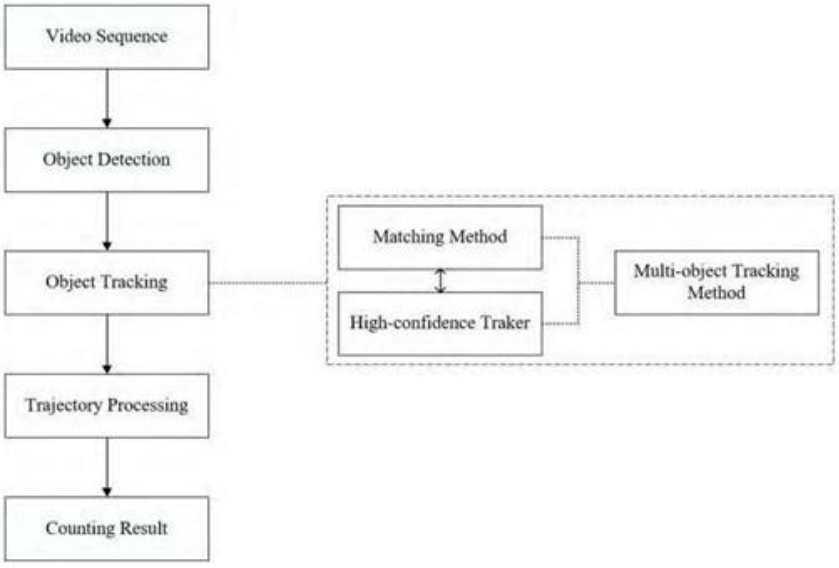


Fig. 1. Vehicle counting framework strategies

The proposed system in the paper[1] presents a video-based approach to vehicle counting, consisting of three primary stages: object detection, object tracking, and trajectory processing. Object detection utilizes the YOLO V3 algorithm, a deep convolutional neural network (DCNN), to identify the position and category of vehicles in each frame of the video. Object tracking combines a novel template matching method with the KCF (Kernelized Correlation Filter) algorithm. Trajectory processing is employed to gather detailed information by clustering trajectories based on their starting point coordinates, utilizing a region encoding model.

The paper [1] introduces two datasets: the Vehicle Counting Dataset (VCD) for validation and the Vehicle Detection Dataset (VDD) used to train the object detection algorithm. The VCD contains videos of urban road and intersection scenes captured by Miovision cameras, while the VDD consists of RGB images annotated with vehicle categories (car, truck, bus).

Major advantage of the proposed system is that it provides enhanced traffic flow information, providing details such as vehicle category, density, and speed.And major disadvantage is that systems may face challenges in handling complex scenes, image blur, noise, and changing environmental conditions, impacting their robustness.

B. YOLO and CLSTM Network Based LOI Counting Model

[2]The proposed framework addresses challenges in counting vehicles in dense traffic scenarios. The existing difficulties include the complex relationship between consecutive frames in dense traffic and the instability introduced by complex tracking methods. To overcome these, a novel spatio-temporal count feature (STCF) is introduced for bi-directional Line of Interest (LOI) counting. The framework uses a YOLO-based detection method to identify salient vehicles, which are then utilized for STCF-based feature extraction. A counting LSTM (cLSTM) network is proposed to establish relationships between continuous frames without relying on complex tracking methods. Previous methods have struggled with dense traffic scenarios, often focusing on individual vehicles. The proposed framework, however, takes a holistic approach, describing dense vehicle flow as a whole. Unlike traditional detection-based methods, the framework emphasizes detecting salient vehicles in a close-up area near the camera.

The proposed methods are evaluated on two datasets: UA-DETRAC and a dataset captured by the lab. The UA-DETRAC dataset includes diverse scenarios such as night, rain, non-straight roads, headlight glare, and off-centered camera positions. It comprises 10 hours of videos recorded at 25 fps with a resolution of 960 × 540 pixels using a Canon EOS 550D camera at 24 locations in Beijing and Tianjin, China, in 2015. Although not originally intended for traffic flow estimation, ten bi-directional traffic videos from UA-DETRAC were chosen for training and evaluation. The lab dataset includes two traffic videos with eight lanes, recorded at 25 fps and aresolution of 960 × 540 pixels using a Canon EOS 70d camera.

Salient vehicle detection involves the initial steps of selecting and initializing a close-up area, followed by the implementation of a YOLO-based model tailored specifically for identifying prominent vehicles. Post-processing procedures, such as probability filtering and bounding box merging, are then applied to enhance the accuracy of the vehicle detection process. Moving on to Spatio-Temporal Counting Feature Extraction (STCF), the method begins with the definition of a crossing probability function for the selected vehicles. Spatial features are extracted, and an STCF is generated for each frame, taking into account the presence of vehicles crossing a predefined counting line.

The cLSTM Network Based Line of Interest (LOI) Counting Model is introduced as the next step. This involves integrating a cLSTM network to analyze bi-directional STCF features, defining output classes based on vehicle crossing directions, and structuring the counting model accordingly. The final stage encompasses the estimation of various traffic flow parameters. This includes the calculation of bi-directional volume based on the counted vehicles, density estimation utilizing the detected vehicles in the close-up area, and speed estimation by considering the relationship between volume and density. This holistic approach combines detection, feature extraction, and counting model design to effectively estimate crucial traffic flow parameters.
Major advantage includes accurate estimation of traffic flow parameters such as speed, volume, and density. And the disadvantage is that this method may not scale well to very large datasets or high-resolution video frames.

C. YOLO and CFNN - Based Intelligent Traffic Monitoring

[3]The research paper proposes an Intelligent Traffic-Monitoring system that leverages the strengths of two different techniques: YOLO (You Only Look Once) for object detection and Convolutional Fuzzy Neural Networks (CFNNs) for vehicle counting and traffic volume estimation. This system aims to address the growing challenges of urban traffic management, including congestion, safety concerns, and inefficient resource allocation. Traditional traffic monitoring methods often rely on manual counting or fixed sensors, which are time-consuming, expensive, and lack real-time adaptability. Inaccuracies and delays in data collection can hinder effective traffic management strategies. As urban populations and vehicle ownership continue to rise, the need for more intelligent and automated solutions becomes increasingly crucial.The proposed system is evaluated using the publicly available UA-DETRAC dataset, which contains 17 video recordings with diverse traffic scenarios and lengths. These videos capture various road conditions, lighting situations, and vehicle types, ensuring the system's robustness in real-world settings.
The system operates in two distinct stages:
Stage 1,harnesses the power of YOLO for efficient object detection. Individual video frames serve as input to YOLO, which meticulously locates and identifies vehicles within each frame. It outputs the detected vehicles along with their corresponding bounding boxes and confidence scores, providing a precise foundation for further analysis.
Stage 2 ,transitions to CFNNs for vehicle counting and traffic volume estimation. It begins by constructing a Time-Space Image (TSI) from the video, a visual representation of traffic flow over a specific timeframe. A CFNN then analyzes the TSI, generating a density map that illuminates the spatial concentration of vehicles. This density map is meticulously parsed to estimate the total vehicle count within the TSI, corresponding to a specific time interval. By processing consecutive density maps, the system effectively estimates the overall traffic volume, expressed in vehicles per hour, for various time intervals.This two-stage approach seamlessly integrates object detection and density map analysis, offering a robust and accurate framework for intelligent traffic monitoring.
The CFNN and Vector-CFNN models proposed in the study not only have favourable vehicle classification effects but also have fewer parameters relative to other models. Therefore, the proposed models are suitable for information analysis in environments with limited hardware performance.

One of the major area of concern about the proposed model is that the complexity of CFNN models, especially with features like the Vector-CFNN model, may result in increased training times, higher computational requirements, and potential challenges in model interpretability.

D. Multi-Regression with Hierarchical Classification and Temporal Regression

[4] proposes a framework which introduces an innovative methodology for vehicle counting on highways through the analysis of compressed video data, a critical aspect of intelligent transportation management systems. Departing from traditional pixel- domain analysis, notorious for its computational intensity, the proposed approach operates directly on compressed video, markedly reducing computational costs. Overcoming challenges associated with limited vehicle information and diverse traffic scenes, the method introduces new low-level features extracted from coding-related metadata, including size, shape, motion, and texture information. These features enhance the accuracy of vehicle counting within the compressed domain. The method incorporates a Hierarchical Classification-based Regression (HCR) model, categorizing traffic scenes based on vehicle density and employing appropriate regression models for estimation. This hierarchical approach effectively captures diverse traffic scene characteristics, significantly improving estimation accuracy. Moreover, this methodology incorporates a localized temporal regression model to enhance counting precision by capitalizing on the ongoing fluctuations in traffic volume over time. Through the fusion of spatial and temporal regression techniques, the suggested approach yields resilient and precise vehicle count outcomes.

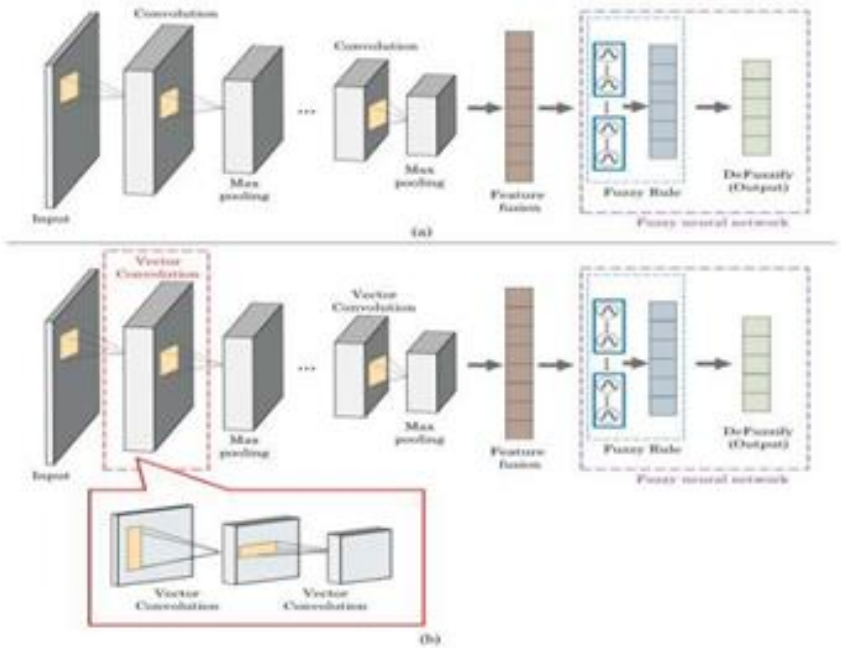


Fig. 2. Strategy of the vehicle counting framework

The structured approach involves four primary phases: video preprocessing, feature extraction, spatial regression-based estima- tion, and temporal regression-based refinement. During preprocessing, metadata from the raw video stream is standardized using techniques like motion vector normalization, macro-block weighting, foreground segmentation, and perspective normalization. Through the HCR model, traffic scenes are categorized into various vehicle density ranges, which enhances estimation accuracy. Additionally, a temporal refinement method is proposed using LOESS, a robust locally weighted regression technique, to

address slow variations in vehicle count from frame to frame. This temporal refinement enhances counting accuracy, acknowledging the typical duration of vehicles within the camera's visual field on a highway. The proposed approach presents a comprehensive solution to the challenges associated with vehicle counting on highways, demonstrating significant advancements in efficiency and accuracy for intelligent transportation systems. The major advantages are the paper introduces a novel method for vehicle count- ing in the compressed domain, addressing a gap in existing research by attempting to achieve comparable performance to pixel-domain methods. The paper proposes new low-level features to mitigate the challenges posed by limited information in compressed videos. These features are easily extracted from coding-related metadata, contributing to the efficiency of the proposed method. The HCR model offers a structured approach by hierarchically dividing traffic scenes based on vehicle density. This helps capture the large variance in traffic scenes, providing a more nuanced understanding of different scenarios.

The disadvantages are that the paper acknowledges that features extracted from compressed videos, such as motion vectors, may be noisy and less accurate compared to raw frames. This could impact the precision of vehicle counting. Relying on encoded metadata might present difficulties since these metadata are primarily optimized for compression efficacy rather than video analysis purposes. This could affect the accuracy of vehicle detection and counting.

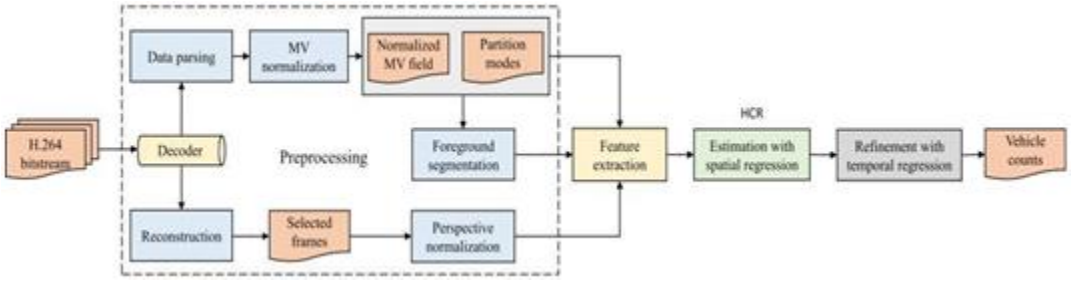


Fig. 3. Framework Of Highway Vehicle Counting System

III. PROPOSED SYSTEM

A. Architecture

The architecture of the traffic control system begins with input data collected from real-time video feeds of traffic intersections. The dataset is then prepared and preprocessed to ensure optimal quality and relevance. Training and test datasets are created for machine learning model development. The ML model is trained to perform object identification, classification, and confidence checking. Upon successful detection, the system proceeds with vehicle type counting, facilitating the traffic management system. Users interact with the system to set priorities for specific vehicle types. Finally, signal control is dynamically adjusted based on real-time data, enhancing traffic flow and efficiency. This comprehensive architecture seamlessly integrates data processing, machine learning, and user interaction for effective traffic control.

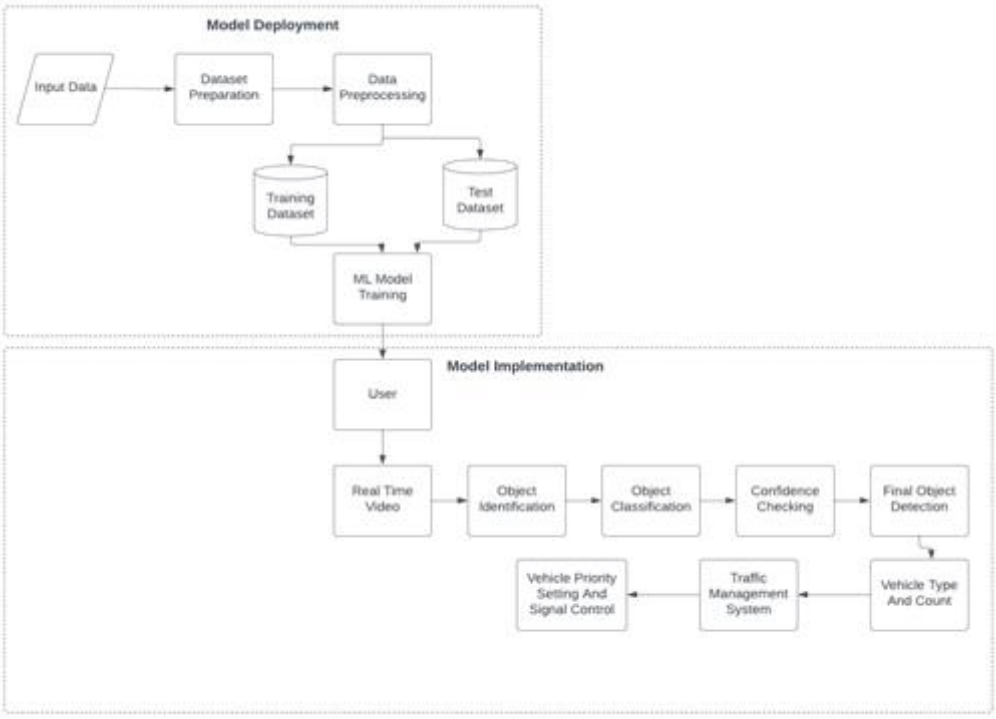


Fig. 4. Architecture of proposed system

B. Implementation

In the proposed system, Video-Based Vehicle detection and Counting for traffic control, we are implementing a Video-Based Vehicle Detection System and Counting for traffic control at junctions. Utilizing YOLO V8 for real-time object detection, our system employs strategically placed high-resolution cameras to accurately identify and classify various types of vehicles passing through the junction. The collected data, including vehicle count and types, is processed by a central unit, and a dynamic traffic control algorithm uses this information to optimize signal timings in real-time. Data collection for a traffic control project using machine learning involves gathering various types of data relevant to traffic flow and behavior. This typically includes; Traffic flow data,Videorecordings,Sensordata,Traffic signal data,Environmentaldata,Weather data and Lighting conditions. Dataset Preprocessing involves removing outliers, missing values, and inconsistent entries that could negatively impact model performance. Techniques like data imputation, outlier detection, and filtering may be employed.This also involves scaling numerical features to a common range, often between 0 and 1, to prevent features with larger scales from dominating the learning process. Techniques like min-max scaling or standard normalization are commonly used.Model Deployment. Once the data has been collected, it needs to be pre-processed before it can be used to train a machine learning model. This preprocessing may involve cleaning the data, removing outliers, and transforming it into a format that the model can understand.The next step is to train a machine learning model on the preprocessed data.Once the model has been trained, it is deployed to a user interface where it can make predictions on new data.Object detection can be used to automatically detect violations such as speeding, red-light running, and illegal parking. This information can help authorities enforce traffic laws and improve safety on the road. YOLO can be used to count and classify vehicles, including cars, trucks, buses, and motorcycles, on a specific road or intersection. This data can be used to measure traffic volume, assess congestion levels, and optimize trafficlight timing. Employing the Traffic Management System

includes integrating YOLO with traffic light system allows for dynamic adjustments based on real-time traffic conditions.By analyzing vehicle count and type, the system can adjust light timings to optimize traffic flow and reduce congestion.

IV. COMPARATIVE STUDY

The Papers were studied, and the comparative results, outlining the advantages and disadvantages of these papers, were compiled into the table below.

TABLE I COMPARATIVE STUDY OF LITERATURE SURVEY

Methods	Advantages	Disadvantages
YOLO V3 and KCF Based Vehicle Counting in Intelligent Transportation	● Versatility ● Low computational cost ● Real Time detection and tracking of objects concurrently	● Sensitivity to Environmental Factors ● Dependency On Video Quality
YOLO and CLSTM Network Based LOI Counting Model	● Accurate estimation of Traffic flow ● Parameters such as speed, volume and density	● Method may not scale large datasets or high resolution video frames.
YOLO and CFNN Based Intelligent Traffic monitoring	● Suitable for information analysis in environments with limited hardware performance.	● Complexity of CFNN models. ● Increased raining times. ● Higher computational requirements, ● Potential challenges in model interpretability.
Multi-Regression with Hierarchical Classification and Temporal Regression	● Innovation in Compressed Domain ● Low level feature Development ● Hierarchial Classification Based Regression(HCR)	● Noisy Compressed Video Features ● Dependence on encoding metadata.

V. CONCLUSION

In conclusion, the evolving landscape of transportation systems demands precision in vehicle counting for the successful implementation of intelligent traffic solutions. Accurate traffic flow information, obtained through video-based vehicle counting systems, serves as a foundational element for crafting effective traffic control strategies. The presented machine vision- based vehicle counting framework, with its tripartite process of detection, tracking, and trajectory processing,showcases the transformative power of deep learning, specifically demonstrated by the deployment of YOLO V8 for robust object detection.Moreover, the practical application of this framework in an intelligent traffic signal control system represents a substantial leap towards responsive

and adaptive traffic management. The capability to dynamically adjust signal timings based on real-time vehicle count data introduces a proactive approach to traffic control, prioritizing intersections with higher vehicle counts. This system holds the potential to significantly enhance traffic flow efficiency, mitigate congestion, and reduce delays in urban environments. The integration of cutting-edge technology not only addresses the limitations of traditional methods but also paves the way for a dynamic and optimized traffic control strategy, marking a notable advancement in the field of intelligent transportation systems.

VI. REFERENCES

[1] Z. Dai et al., ”Video-Based Vehicle Counting Framework,” in IEEE Access, vol. 7, pp. 64460-64470, 2019, doi: 10.1109/ACCESS.2019.2914254.

[2] ”BI-Directional Dense Traffic Counting based on Spatio-Temporal Counting Feature and Counting-LSTM(2021)”Shuang Li , Faliang Chang , and Chunsheng Liu , Member, IEEE.

[3] ”Intelligent Traffic - Monitoring System Based on YOLO and Convolutional Fuzzy Neural Networks .”(2022) Cheng-Jian Lin(Senior Member,IEEE), and Jyun-Yu Jhang

[4] ”Compressed-Domain Highway Vehicle Counting by Spatial and Temporal Regression.”

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

AI Enhanced Recruitment

Ms. Remya Paul*[1], Abhinav P George[2], Alvin Saju[2], Anto S Illickal[2], Sharath Sivakumar[2]
*[1]Assistant Professor, Department of Computer Science and Engineering, VJCET, Ernakulam, Kerala, India
[2]Student, Department of Computer Science and Engineering, VJCET, Ernakulam, Kerala, India

ABSTRACT
The AI-Enhanced Recruitment Platform revolutionizes talent acquisition by harnessing cutting-edge artificial intelligence technologies. Through sophisticated natural language processing (NLP) and machine learning (ML) algorithms, the platform intelligently sources candidates from diverse channels, extracting pertinent skills and experiences to match specific job requirements. Automated screening processes swiftly analyze candidate profiles, alleviating the burden on recruiters and enabling them to focus on building relationships with the most promising candidates. Personalized candidate matching further enhances engagement and retention by ensuring tailored job recommendations based on individual skills, experience, and cultural fit. Notably, the platform incorporates unbiased assessment tools, minimizing unconscious bias and fostering fair hiring practices. This approach leads to numerous benefits, including reduced time to hire, improved candidate experiences, decreased unconscious bias, and enhanced cost-effectiveness through streamlined processes and lowered recruitment expenses. Moreover, AI-powered analytics provide invaluable insights into candidate pools, market trends, and hiring effectiveness, facilitating data-driven decision-making and maintaining a competitive edge in talent acquisition. In summary, the AI-Enhanced Recruitment Platform represents a pivotal advancement in the recruitment landscape, offering organizations a transformative solution to efficiently attract and retain top talent.

Keywords: Keyword Extraction, CNN, Machine Learning, Natural Language Processing,

I. INTRODUCTION

The AI-Enhanced Recruitment Platform represents aholistic approach to talent acquisition, driven by advancednatural language processing (NLP) and machine learning(ML) algorithms. One of its hallmark features is intelligentsourcing, which enables organizations to comb throughvast volumes of resumes, profiles, and online sources withunparalleled accuracy. By extracting relevant skills andexperiences, the platform effectively identifies andqualifies candidates who align with specific jobrequirements, thereby streamlining the initial stages of therecruitment process.Furthermore, the platform automated screening capabilities significantly reduce the burden on recruiters by employing ML algorithms to analyze candidate profiles against job criteria. This automation not only expedites candidate evaluation but also ensures a more objective and consistent screening process. As recruiters are freed from repetitive tasks, they can allocate more time and resources

towards building meaningful relationships with the most promising candidates, thereby enhancing the overall candidate experience.

A cornerstone of the AI-Enhanced Recruitment Platform is its commitment to fairness and inclusivity. By incorporating unbiased assessment tools powered by AI, the platform minimizes the impact of unconscious bias in decision-making processes. This not only promotes diversity and equity in hiring practices but also ensures that the best candidates are selected based on merit and qualification alone.Moreover, the platform delivers a myriad of benefits beyond operational efficiencies. Reduced time-to-hire, improved candidate experiences, decreased unconscious bias, and enhanced cost-effectiveness are among the tangible advantages organizations can expect to gain. Additionally, the platform provides invaluable data-driven insights into candidate pools, job market trends, and recruitment effectiveness, enabling organizations to make informed decisions and gain a competitive edge in talent acquisition.

II. LITERATURE SURVEY

A. AUTOMATIC KEYWORD EXTRACTION

The presented method for keyword extraction offers a meticulous three-stage approach, furnishing a comprehensive framework to derive meaningful keywords from processed documents. The initial phase involves the judicious selection of candidate keywords, where the document undergoes preprocessing through segmentation, tokenization, and Parts of Speech (PoS) tagging. Notably, common English stopwords are excluded, and a regular matching procedure is implemented to extract noun chunks, predominantly comprising nouns and adjectives. Further refinement is achieved through length filtering, eliminating candidate keywords exceeding five words. This meticulous curation ensures a focused pool of potential keywords, enhancing the subsequent stages of the method.

The second stage revolves around word scoring, a pivotal step in keyword extraction. The TextRank algorithm is instrumental, treating the text as a graph where nodes represent words and edges signify co-occurrence within a sliding window. The algorithm assigns scores to words based on their importance within the interconnected graph. Simultaneously, semantic relationships are incorporated through methods such as Word Embedding, WordNet, and Normalized Google Distance. These diverse approaches enable the method to capture both syntactic and semantic nuances, overcoming limitations associated with singular methodologies. The innovation lies in the introduction of three distinct combinations, intelligently amalgamating co-occurrence and semantic relationships. This sophisticated integration enriches the evaluation of word importance, providing a nuanced perspective on the relevance of words within the document.

The final stage of keyword extraction involves ranking candidate keywords based on their scores, eliminating duplicates, and selecting the top K keywords. The methodology's effectiveness is substantiated through an experimental evaluation, comparing the quantity and quality of selected candidate keywords before and after the application of the proposed method. This empirical validation underscores the method's prowess in optimizing keyword extraction. By adeptly combining syntactic and semantic features, the method not only enhances the precision of keyword selection but also contributes to a deeper understanding of the document's content. The seamless integration of these stages presents a robust and adaptable approach, poised to elevate the efficiency of keyword extraction across diverse domains and document types.

B. TEXTRANK MODEL

The methodology employed in the study utilizes the TextRank algorithm for keyword extraction, leveraging a text graph representation where words are nodes and edges denote co-occurrence relationships within a specified window size. This approach involves several key steps. First, the text graph is constructed as an undirected graph, with words as nodes and edges representing co-occurrence relationships within a given window size. Words appearing within this window are treated as connected nodes, forming the basis of the graph. TextRank scores are then calculated for each node in the text graph using the PageRank algorithm. These scores determine the importance of each word within the text, and the top-k words with the highest TextRank scores are selected as keywords.

The study conducts experiments on two datasets, Hulth2003 and Krapivin2009, which are commonly used for keyword extraction tasks. Evaluation metrics such as Precision, Recall, and Accuracy are employed to assess the performance of TextRank. Various parameters including co-occurrence window size (w), iteration number (t), decay factor (c), rank (k), and text length are tested to analyze their impact on TextRank's performance. The experimental results reveal insights into the effect of different parameters on TextRank's effectiveness. For instance, Precision initially increases with the co-occurrence window size, stabilizing or decreasing thereafter, with an optimal window size suggested at 3. Similarly, the iteration number influences Precision, with a recommended setting of 20 for stability. Additionally, the study finds that the decay factor and rank in returned lists also affect TextRank's performance.

The study provides valuable insights into optimizing TextRank for keyword extraction tasks, highlighting optimal parameter settings based on experimental results. Furthermore, the robustness of TextRank across different settings and datasets is demonstrated, with its performance appearing independent of text length. The study references foundational algorithms such as PageRank and describes the co-occurrence relation used to build the text graph. Moreover, it cites previous works that have used the Hulth2003 dataset for keyword extraction evaluation, while also mentioning the evaluation metrics employed including Precision, Recall, and Accuracy.

C. FACIAL EXPRESSION RECOGINITION

The proposed model encompasses a holistic framework for facial expression recognition (FER), face synthesis, and face alignment, depicted in Figure 2. Comprising three modules—face alignment network (green), face synthesis network (orange), and FER network (blue)—the model seamlessly integrates geometry information, expression codes, and identity representation. The face alignment network predicts landmarks used in face synthesis and as geometry information for FER. The face synthesis module utilizes an encoder-decoder network to generate labeled facial images under arbitrary poses. Adversarial learning, content-similarity, and perceptual losses ensure accurate synthesis while preserving identity and expression. The FER network, derived from weighted fusion of appearance and geometry features, attains joint FER, face synthesis, and face alignment. The unified learning objectives encompass minimizing Mean Square Error (MSE) for landmark prediction, adversarial losses for face synthesis, and softmax cross-entropy loss for FER.

In the learning phase, objectives include training the face alignment network, synthesizing facial images underground truth landmarks, alongside a classification loss enforcing landmark constraints. Face synthesis employs adversarial learning, content-similarity, and perceptual losses, ensuring identity preservation and expression synthesis. The FER network combines appearance and geometry features, facilitating joint learning of expression, geometry, and synthesized faces. The model demonstrates its versatility by simultaneously

addressing FER, face synthesis, and face alignment challenges, establishing a unified framework for enhanced performance.

Comparative analysis with existing works highlights distinctions from geometry-contrastive GAN (GC-GAN), joint AU detection and face alignment framework (JAA-Net), and conditional GAN-based FER (CG-FER). While sharing similarities with CG-FER in GAN-based face synthesis for FER improvement, the proposed method distinguishes itself through flexible geometry representation, additional perceptual loss, and the incorporation of face alignment. This unified approach leverages joint relationships among expression, landmarks, and synthesized faces, setting it apart from existing models and enhancing performance across diverse facial analysis tasks.

D. 3D FACIAL TRACKING

The provided text outlines a comprehensive 3D facial tracking framework with a focus on a parametric face model and the GoMBF-Cascade motion regression method. The parametric face model employs a 3D facial mesh representation, incorporating linear basis vectors for identity and delta blendshapes for expressions. The tracking workflow involves initializing facial motion parameters, training the GoMBF-Cascade through progressive refinement, and globally optimizing the model for improved accuracy. This framework combines a modular boosted ferns approach for training and global optimization of the model's leaves to achieve an effective representation of 3D facial motion.

In the GoMBF-Cascade regression process, training involves the creation of guess-truth pairs, modular learning for individual motion parameter categories, and globaL optimization for enhanced performance. During runtime prediction, the Viola-Jones detector and SDM model are employed for face and landmark detection. The parametric face model is fitted to detected 2D landmarks to estimate initial motion parameters, followed by iterative refinement through the GoMBF-Cascade approach. Appearance vector extraction involves generating feature points around reference 2D landmarks, indexing based on barycentric coordinates, and extracting pixels from the image, providing a discriminative representation of facial motion. This framework aims to robustly track 3D facial motion in monocular RGB videos, demonstrating a combination of a parametric face model and a cascaded regression approach for improved tracking accuracy.

III. PROPOSED SYSTEM

A. ARCHITECTURE

The proposed AI-Enhanced Recruitment Platform is poised to revolutionize talent acquisition through a streamlined workflow. Commencing with a sophisticated resume screening process utilizing advanced keyword extraction, the platform accelerates candidate shortlisting, optimizing recruiters' time. Shortlisted candidates undergo a rigorous online skills test, ensuring a thorough evaluation of their capabilities and validating qualifications. The system culminates in an AI-powered behavioral interview, leveraging artificial intelligence to analyzebehavioral nuances, communication skills, and relevant traits, facilitating more informed decision-making. This end to-end approach offers a comprehensive evaluation of candidates, emphasizing efficiency and precision in talent acquisition. By strategically integrating AI at multiple stages, the platform aims to redefine how organizations identify and engage top-tier talent, providing a modern solution for the evolving needs of recruitment in a competitive job market. In summary, the proposed system combines advanced technology

with a holistic assessment approach, promising to be a game-changer in enhancing recruitment strategy and transforming talent acquisition processes.The architecture of the AI-Enhanced Recruitment Platform is designed with scalability and adaptability in mind. It leverages cloud computing infrastructure to handle vast amounts of data efficiently, ensuring seamless operation even during peak recruitment periods. The modular design allows for easy integration with existing HR systems, minimizing disruption to established workflows. Additionally, the platform employs robust security measures, including encryption and access controls, to safeguard sensitive candidate information. Continuous monitoring and updates ensure compliance with evolving data privacy regulations, instilling trust and confidence in both candidates and hiring organizations. With its agile architecture and commitment to data integrity, the platform serves as a reliable foundation for driving innovation in talent acquisition practices.

B. IMPLEMENTATION

The Resume Screeningstage begins with the input of resumes from job applicants. The process involves the application of sophisticated screening techniques, utilizing keyword extraction to efficiently identify crucial qualifications. The output of this stage is a list of shortlisted candidates who match the specified criteria. This streamlined approach accelerates the candidate selection process, saving valuable time for recruiters and ensuring that only the most qualified individuals progress to the next stages of evaluation. The emphasis on keyword extraction enhances the precision of the screening process, aligning the initial evaluation with the specific requirements of the job. This phase serves as a crucial initial filter, setting the foundation for a more focused and efficient talent acquisition process.The Online Skills Test phase begins with the input of shortlisted candidates identified during the resume screening. In this stage, candidates undergo a tailored online skills test designed to assess their capabilities in alignment with the job requirements. The output of this process is the\ identification of skill-validated candidates who have demonstrated proficiency in the key areas essential for the role. This phase serves as a practical and in-depth evaluation, ensuring that candidates not only possess the qualifications highlighted in their resumes but also have the specific skills required for the job. By incorporating a skills test into the screening process, the platform adds a layer of practical validation, enhancing the overall precision of candidate evaluation and contributing to a more thorough and informed decision-making process in subsequent stages of recruitment.The AI-Powered Behavioral Interviewphase begins with skill-validated candidates identified in the previous stages. During this stage, candidates undergo a behavioral interview conducted by artificial intelligence. The process involves assessing various behavioral aspects, communication skills, and relevant traits. The output of this phase is a pool of interview evaluated candidates, providing recruiters with valuable insights into the candidates' inter personal skills, problem-solving abilities, and cultural fit within the organization. By lever aging AI for the interview process, the platform introduces an objective and standardized approach, ensuring a consistent evaluation across all candidates. This data-driven evaluation contributes to more informed decision-making, offering a comprehensive understanding of each candidate's suitability for the role beyond their technical skills. The behavioral interview phase represents a forward-thinking and efficient method for gauging the holistic qualities of candidates, aligning with the platform's commitment to redefining talent acquisition.The Decision Makingphase takes input fromcandidates who have undergone the AI-poweredbehavioral interview. In this stage, recruitersanalyse thecomprehensive evaluation results tomake informed decisions. The process involvesconsidering various factors, including technicalskills, behavioral aspects, and cultural fit. Theoutput of this phase is the selection ofcandidatesdeemed

most suitable for the role, advancing them for further engagement in the recruitment process.Recruiters rely on the data-driven insights obtainedfrom the behavioral interviews to ensure a well-rounded understanding of each candidate'scapabilities and potential contribution to theorganization. Thisstage emphasizes the platform'scommitment to precision in talent acquisition,enabling recruiters to make strategic and informeddecisions that align with the organization's goals.The selected candidates proceed to the final stages,facilitating a seamless transition from evaluation toengagement in the recruitment process.

IV. COMPARATIVE STUDY

TABLE I COMPARITIVE STUDY OF LITERATURE SURVEY

Paper	*Advantage*	*Disadvantage*
Automatic Keywords Extraction based on Co-occurrence and Semantic Relationships between Words	Achieves good performance on sentiment classification tasks, demonstrating its ability to capture both keywords and sentiment information from text..	Dependence on specific sentiment lexicons and the TextRank algorithm may make it prone to overfitting, particularly when applied to new domains or data distributions.
An Empirical Study of Textrank for Keyword Extraction	Demonstrates that TextRank is a robust and effective method for keyword extraction, achieving good performance across different datasets and parameter settings.	Focuses exclusively on TextRank and does not compare it to other keyword extraction methods.
A Unified Deep Model for Joint Facial Expression Recognition, Face Synthesis, and Face Alignment	Applied to a wide range of applications, including human-computer interaction, affective computing, and virtual reality.	The deep learning nature of the model makes it difficult to interpret its internal workings and understand how it arrives at its decisions
Real-Time 3D Facial Tracking via Cascaded Compositional Learning	The system achieves high accuracy in 3D gaze estimation, even under various lighting conditions, poses, shapes, and facial expressions	. The evaluation is primarily conducted on live videos and Internet videos, and further evaluation on more diverse datasets could strengthen the conclusions.

V. CONCLUSION

In conclusion, the "AI-Enhanced Recruitment Platform" presents a paradigm shift in talent acquisition, promising to redefine traditional recruitment methods. Its innovative workflow, beginning with a sophisticated resume screening process utilizing keyword extraction, swiftly identifies crucial qualifications, streamlining candidate shortlisting and saving invaluable time for recruiters. The inclusion of an online skills test ensures a thorough assessment of candidates' capabilities, adding depth to the evaluation process. However, the true game-changer lies in the platform's introduction of an AI-powered behavioral interview, which not only enhances efficiency but brings unprecedented precision to talent acquisition. This comprehensive approach, analyzing candidates across various dimensions, marks a departure from conventional practices. Organizations are urged to elevate their recruitment strategy with this transformative platform, signaling a shift towards modernity and strategic precision in identifying and engaging top-tier talent. As the recruitment landscape evolves, the AI-Enhanced Recruitment Platform stands at the forefront, offering not just time savings but a progressive and data-driven solution for making informed decisions in a competitive job market. In embracing this platform, organizations position themselves to navigate the complexities of talent acquisition with agility, ensuring a more streamlined, informed, and effective recruitment process that aligns with the dynamic needs of the contemporary workforce.

VI. REFERENCES

[1] XIANGKE MAO, SHAOBIN HUANG , RONGSHENG LI, AND LINSHAN SHEN "AUTOMATICKEYWORDS EXTRACTION BASED ON CO-OCCURENCE AND SEMANTIC RELATIONSHIPS BETWEEN WORDS"College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China}

[2] MINGXI ZHANG, XUEMIN LI, SHUIBO YUE, AND LIUQIAN YANG "AN EMPIRICAL STUDY OF TEXTRANK FOR KEYWORD EXTRACTION"Grant No. 16ZR1422800 vol. 3, pp. 993–1022, 2003.

[3] Feifei Zhang, Tianzhu Zhang, Member, IEEE, Qirong Mao, Member, IEEE, and Changsheng Xu "A UNIFIED DEEP MODEL FOR JOINT FACIAL EXPRESSION,FACE SYNTHESIS AND FACE ALIGNMENT" 10.1109/TIP.2020.2991549, IEEE transactions on Image Processing

[4] JianwenLou ,Xiaoxu Cai, Junyu Dong , Member, IEEE, and Hui Yu , Senior Member, IEEE"REAL-TIME 3D FACIAL TRACKING VIA CASCADED COMPOSITIONAL LEARNING"IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 30, 2021

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

Scratch Detection on Vehicles

Ms. Anu Jose*[1], Anannya Mathew[2], Gadha Ashok[2], Nehala Kassim[2]
*[1]Assistant Professor, Department of CSE, VJCET Ernakulam, Kerala, India
[2]Student, Department of CSE, VJCET, Ernakulam, Kerala, India

ABSTRACT
The vehicle scratch detection system utilizes deep learning, specifically convolutional neural networks (CNNs), to analyze high-resolution images captured by vehicle-mounted cameras or sensors. Through the CNN architecture, the system learns intricate patterns and features that distinguish scratches from other elements on the vehicle's body. Training the CNN involves a substantial dataset of labeled images, where each image indicates the presence or absence of scratches. Techniques like data augmentation expand the dataset, enabling the model to generalize better and recognize diverse scratch patterns. Once trained, the CNN can swiftly analyze incoming images in real-time scenarios, revolutionizing vehicle maintenance by promptly detecting scratches. This capability has significant potential in automating inspection and repair tasks, enhancing vehicle aesthetics, and improving customer satisfaction by maintaining the vehicle's pristine appearance.

Keywords: RCNN, CNN, YOLO, FCNN, FPT, PCA

I. INTRODUCTION

The landscape of scratch detection in automotive applications has seen significant advancements through the integration of deep learning and image segmentation techniques. Recent studies have focused on developing more accurate, efficient, and nuanced methods for identifying and assessing scratches on various vehicle parts, from stamped automotive components to micro-scratches that are challenging to detect with conventional methods.

[1] introduces a dual-phase approach where deep learning models first identify potential scratch areas, followed by precise segmentation to delineate the scratches accurately This method leverages the strengths of convolutional neural networks (CNNs) to handle the initial identification task, ensuring high sensitivity to scratches of varying sizes and severities, and then applies advanced image segmentation algorithms for fine-grained analysis.

[2] focuses specifically on the challenges presented by the reflective surfaces of stamped parts. The authors propose a specialized framework that combines adaptive lighting techniques with a tailored CNN architecture to enhance the visibility of scratches under variable lighting conditions, significantly improving detection accuracy and consistency

[3]the emphasis is on detecting micro-scratches that are typically overlooked by standard inspection systems. By employing a deep CNN model trained on a meticulously annotated dataset of high-resolution images, the study demonstrates the feasibility of detecting even the most subtle surface imperfections, paving the way for enhanced quality control in precision manufacturing

[4] expands the scope of scratch detection to encompass broader vehicle damage assessment. Utilizing an improved Mask R-CNN framework, this research addresses both the detection and accurate segmentation of various types of vehicle damage, including scratches, dents, and more substantial structural damages. The enhanced algorithm offers improved speed and accuracy, facilitating real-time damage assessment in practical scenarios (Author D, 2023).

Collectively, these studies underscore the pivotal role of deep learning and image segmentation in revolutionizing scratch and damage detection in the automotive industry. By tailoring algorithms to specific challenges—be it the detection of fine micro-scratches or the segmentation of damage on complex surfaces—researchers are setting new benchmarks for accuracy, efficiency, and applicability in real-world settings.

II. LITERATURE SURVEY

A. DEEP LEARNING

Single-shot multibox detector (SSD) is a real-time object detection algorithm that can detect multiple objects in an image with a single forward pass through the network. It uses a convolutional neural network (CNN) to extract features from the image and then predicts bounding boxes and class labels for each object. SSD is faster than other object detection algorithms like Faster R-CNN, but it may not be accurate. Up sampling-lateral-fusion (ULF) is a module used in the SIL-Net architecture to improve the accuracy of object detection. It up samples the feature maps from the lower layers of the network and fuses them with the feature maps from the higher layers. This helps to improve the localization and classification accuracy of the detector. Principal Component Growth Segmentation (PGS) algorithm is a novel algorithm for scratch segmentation used in image processing. It starts with a small set of seed points and then grows them by adding new points that are similar to the existing points. This process continues until the desired segmentation is achieved. Principal Component Analysis (PCA) is a statistical technique used for dimensionality reduction and feature extraction. It identifies the principal components of a dataset, which are the directions of greatest variance in the data. By projecting the data onto the principal components, it can reduce the dimensionality of the data without losing much information.

B. NRPCA

Traditional image-based inspection methods for automotive stamping parts face several limitations like limited depth information, sensitivity to lighting, complex part geometry.3D point cloud technology offers solutions to these challenges like rich surface representation, lighting independence and detailed feature extract. By utilizing Point Cloud Geometric Characteristics for Scratch Detection proposes a novel scratch detection method based on analyzing the geometric characteristics of 3D point clouds by Point Cloud Preprocessing which is Denoising and filtering the point cloud to remove outliers and improve data quality,Local Feature Extraction which is Computing various local geometric features for each point, such as height, curvature, and normal vectors.Scratch Candidate Identification which Identifies potential scratch regions based on specific feature thresholding or statistical analysis and False Positive Reduction which

Employs additional feature analysis or machine learning algorithms to eliminate false positives and refine the scratch detection results.

C. CNN

Scratching a car might seem like a minor inconvenience, but in the demanding world of automotive manufacturing, even the faintest imperfections on metal surfaces can be a major problem.Low Contrast, Varied Illumination,Complex Surface Geometry,Noise and Imperfections: The natural texture of metal surfaces, along with dust, grease, or other contaminants, can introduce noise and clutter that complicate scratch detection algorithms these challenges often lead to missed defects, impacting product quality, increasing rework costs, and potentially compromising safety. To overcome limitations of traditional method researchers are turning to deep convolutional neural networks (DCNNs). DCNNs excel at extracting complex features from images and learning to distinguish subtle patterns .A DCNN method could work by Training Data,Network Architecture,Feature Learning,Scratch Detection.

D. RCNN

Feature Extraction with ResNet50+FPN works by Input Image: An image is fed into the model. A pre-trained ResNet50 convolutional neural network extracts deep features from the image at multiple scales. Feature Pyramid Network (FPN) combines these features to create multi-scale feature maps, preserving both high-level semantic information and fine-grained details. Region Proposal Generation with RPN scans the feature maps at multiple scales, At each window position, multiple pre-defined anchor boxes of different sizes and aspect ratios are considered. The RPN classifies each anchor box as either foreground (likely containing an object) or background. For foreground anchors, the RPN refines their positions to better match potential object boundaries. MS filters out overlapping candidate boxes, retaining the most promising ones for further processing. RoIAlign for Precise Alignment feature maps and remaining ROIs are passed to the RoIAlign layer. It extracts fixed-size feature maps for each ROI, ensuring precise spatial alignment between the ROIs and the extracted features. Finally, Splitting for Object Detection and Segmentation ss done

III. PROPOSED SYSTEM

The proposed vehicle scratch detection system using deep learning revolves around a convolutional neural network (CNN) architecture tailored for precise scratch identification. It involves assembling a diverse, labeled dataset of high-resolution images capturing various scratch types and non-damaged surfaces. Through iterative training and validation, the CNN learns to differentiate between scratches and non-scratch areas, potentially employing data augmentation techniques for robustness. Once trained, the system will be implemented to process real-time images captured by onboard sensors or cameras, swiftly and accurately detecting scratches. Rigorous testing and evaluation will ensure its reliability across different environmental conditions and scratch variations. Ultimately, the system aims to integrate seamlessly into vehicle maintenance infrastructure, offering an automated and efficient solution for prompt scratch identification and facilitating timely repairs, thereby enhancing vehicle aesthetics and customer satisfaction.

A. Architecture

The architecture for scratch detection is designed to employ advanced computer vision techniques to identify and locate scratches on surfaces, particularly on vehicles. Beginning with the acquisition of images through cameras or sensors, the process involves preprocessing steps like resizing, normalization, and noise reduction to enhance image quality. Feature extraction techniques are then applied to highlight relevant aspects indicative of scratches, followed by the implementation of a machine learning or deep learning model. This model is trained using labeled datasets, allowing it to learn and recognize patterns associated with scratches. During the inference stage, the trained model is applied to new images, generating predictions about the presence and location of scratches. Post-processing steps may refine these results, and the final outcomes are often visualized and integrated into alert systems for real-time applications. The architecture ensures a systematic and accurate approach to detecting scratches on vehicle surfaces.

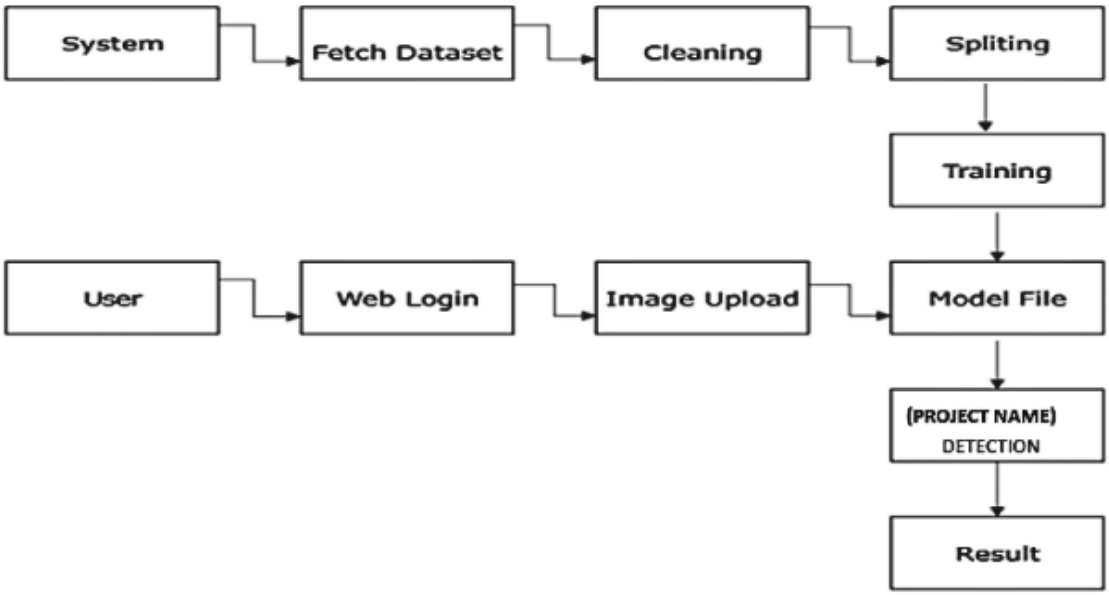


Fig-1: Architecture (YOLO)

B. Implementation

The implementation of vehicle scratch detection is a fascinating journey into the realm of artificial intelligence, transforming it into an eagle-eyed partner dedicated to safeguarding your car's value and appearance. Commencing with the assembly of a comprehensive image library, vehicles of various shapes and sizes showcase their battle scars and pristine surfaces under diverse lighting conditions. This library becomes a trove of scratchy intel, refined meticulously to filter out blurry photos, mislabeled marks, and irrelevant clutter, ensuring the AI learns from accurate representations.

The library is then divided into training and testing sets. The training set becomes a battleground where the AI learns to identify scratches with laser-like precision, while the testing set acts as a silent observer, assessing the AI's ability to translate its skills beyond the training arena. The real magic unfolds with deep learning – intricate neural networks, acting as seasoned detectives, scrutinize millions of images. They

discern subtle lines, color variations, and textures that betray the presence of a scratch, evolving into masters of detecting even the faintest blemish.

With training complete, the AI warrior is unleashed. When a new image enters the scene, the AI swiftly scans every pixel, analyzing details with lightning speed to identify scratches. The boundaries of the scratch are precisely mapped, unleashing endless possibilities. Insurance companies receive instant damage reports, automated repair bots target identified scratches with surgical precision, and car washes optimize cleaning processes based on detected blemishes. Scratches, once mere annoyances, now become whispers revealing stories, guiding repairs, and ensuring the safety and value of the vehicle.

This implementation showcases the transformative power of AI in the automotive world. The next time a tiny mark appears on your car, remember that AI stands ready as your scratch sleuth, ensuring your vehicle looks and performs at its best. Looking ahead, the mastery of deciphering the hidden language of scratches by AI detectives may even contribute to the navigation of self-driving cars in tight spaces or avoidance of potential hazards, marking a new era in automotive technology.

REFERENCE	TITLE	TECHNIQUE	MERITS	DEMERITS
[1]	A Scratch Detection Method Based on Deep Learning and Image Segmentation	• Deep Learning • Image Segmentation • PGS	• Scratch Segmentation • High Accuracy	• Complex Algorithm • Limited Adaptability
[2]	A Novel Scratch Detection and Measurement Method for Automotive Stamping Parts	• BWI • NRPCA • P-NRPCA	• Improved Accuracy with 3-D Data • Reduced Environmental Dependence	• Limited Generalizability • Complexity
[3]	Weak Micro-Scratch Detection Based on Deep Convolutional Neural Network	• CNN • Du-Net	• Efficiency • Non-Destructiveness	• Expert Knowledge • Computational Resources
[4]	Vehicle-Damage-Detection Segmentation Algorithm Based on Improved Mask RCNN	• Improved Mask RCNN • Res-Net • SSD	• High Accuracy • Instance Segmentation	• Complexity • Large Training Data Requirements

Table1:Comparative Study

IV. CONCLUSION

In conclusion, the integration of deep learning, specifically utilizing the robust YOLO architecture, in the context of a vehicle scratch detection system represents a significant shift in the automotive maintenance landscape. YOLO's proficiency in object detection, coupled with its incorporation into deep learning

approaches, enables the system to quickly and accurately identify scratches across a vehicle's surface. Through extensive training on meticulously annotated datasets that encompass various scratch variations, YOLO excels in rapidly identifying and categorizing these imperfections with exceptional precision. When seamlessly integrated into on board vehicle systems, this technology operates autonomously, swiftly detecting and flagging scratches during routine operations, optimizing maintenance procedures, and expediting repairs. This ground breaking advancement not only enhances vehicle aesthetics and customer satisfaction but also transforms automotive maintenance, fundamentally enhancing efficiency by promptly addressing surface imperfections, thereby ensuring the longevity and visual appeal of vehicles.

V. REFERENCES

[1] A Scratch Detection Method Based on Deep Learning and Image Segmentation (LemiaoYang , Fuqiang Zhou , and Lin Wang)

[2] Liu, B. An, Y. Hou, H. Wang and Y. Liu, "A Novel Scratch Detection and Measurement Method for Automotive Stamping Parts," in IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1-13, 2022, Art no. 5017213, doi: 10.1109/TIM.2022.3193970.

[3] L. Song, W. Lin, Y. -G. Yang, X. Zhu, Q. Guo and J. Xi, "Weak Micro-Scratch Detection Based on Deep Convolutional Neural Network," in IEEE Access, vol. 7, pp. 27547-27554, 2019, doi: 10.1109/ACCESS.2019.2894863.

[4] Q. Zhang, X. Chang and S. B. Bian, "Vehicle-Damage-Detection Segmentation Algorithm Based on Improved Mask RCNN," in IEEE Access, vol. 8, pp. 6997-7004, 2020, doi: 10.1109/ACCESS.2020.2964055.

[5] H. Zheng, L. Fang, M. Ji, M. Strese, Y. Özer and E. Steinbach, "Deep Learning for Surface Material Classification Using Haptic and Visual Information," in IEEE Transactions on Multimedia, vol. 18, no. 12, pp. 2407-2416, Dec. 2016, doi: 10.1109/TMM.2016.2598140.

[6] K. Han, M. Sun, X. Zhou, G. Zhang, H. Dang and Z. Liu, "A new method in wheel hub surface defect detection: Object detection algorithm based on deep learning,"2017 International Conference on Advanced Mechatronic Systems (ICAMechS) Xiamen, China, 2017, pp. 335-338, doi: 10.1109/ICAMechS.

[7] J. Sun, P. Wang, Y. -K. Luo and W. Li, "Surface Defects Detection Based on Adaptive Multiscale Image Collection and Convolutional Neural Networks," in IEEETransactions on Instrumentation and Measurement, vol. 68, no. 12, pp. 4787-4797, Dec.2019, doi: 10.1109/TIM.2019.2899478.

[8] R. Borwankar and R. Ludwig, "A Novel Compact Convolutional Neural Network for Real-Time Non destructive Evaluation of Metallic Surfaces," in IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 10, pp. 8466-8473,Oct. 2020, doi: 10.1109/TIM.2020.2990541.

[9] R. Usamentiaga, D. G. Lema, O. D. Pedrayes and D. F. Garcia, "Automated Surface Defect Detection in Metals: A Comparative Review of Object Detection and Semantic Segmentation Using Deep Learning," in IEEE Transactions onIndustryApplications,vol.58,no.3,pp.4203-4213,MayJune 2022, doi:10.1109/TIA.2022.315156

[10] L. Renwei and Y. Dong, "Component surface defect detection based on image segmentation method," 2016 Chinese Control and Decision Conference (CCDC), Yinchuan, China, 2016, pp. 5093-5096, doi: 10.1109/CCDC.2016.7531906.

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

Transformative Advances in Medical Coding: Introducing Medcode, A Neural Network Approach

Ms Lakshmi Suresh*[1], Anna Prize Johney[2], Aysha Nazrin Afsal[2], Lekshmipriya C M[2]
*[1]Assistant Professor, Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Muvattupuzha, Kerala, India
[2]Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Muvattupuzha, Kerala, India

ABSTRACT
This project IoT based Coal mine safety monitoring and alerting system based on sensor network can timely and accurately reflect situation of environment and staff in the underground regions to computer system. The air pollution from coal mines is mainly due to emissions of hazardous gases include Sulphur dioxide (SO2), nitrogen dioxide (NO2), carbon monoxide (CO) etc. To inside the mine. This system also provides an early warning, which will be helpful to all miners present inside the mine to save their life before any accidents occurss

Keywords : Coal Mines, ESP 32 Camera, Coal Mines, Sensors

I. INTRODUCTION

In the ever-evolving landscape of healthcare, the accurate translation of complex medical information into standardized codes is paramount for ensuring seamless communication among healthcare professionals, insurers, and regulatory bodies. This process, known as medical coding, plays a pivotal role in facilitating reimbursement, streamlining administrative tasks, and ultimately improving patient care. As the volume and diversity of healthcare data continue to burgeon, the demand for efficient and precise medical coding solutions becomes increasingly imperative. This research embarks on a transformative journey by delving into the realm of Natural Language Processing (NLP) to redefine and enhance medical coding processes through the development of a cutting-edge neural network architecture termed MEDCODE (Medical Coding Neural Network). Medical coding involves assigning alphanumeric codes to various medical procedures, diagnoses, and services. These codes, typically derived from universally accepted coding systems such as the International Classification of Diseases (ICD) and Current Procedural Terminology (CPT), serve as a standardized language that facilitates accurate documentation and billing across the healthcare ecosystem. The intricacies of medical coding, however, are compounded by the vast array of clinical documentation, including electronic health records, clinical notes, and medical reports. The complexity of healthcare narratives demands a sophisticated

solution that transcends traditional coding approaches and harnesses the power of advanced computational techniques.

Against this backdrop, our research endeavors to bridge the gap between the intricacies of medical language and the need for precise coding through the implementation of state-of-the-art NLP methodologies. The proposed MEDCODE system represents a pioneering effort to leverage neural network architectures in the domain of medical coding, with a focus on not only semantic understanding but also the nuanced intricacies specific to healthcare documentation.

The overarching goal of the MEDCODE system is to streamline and enhance the medical coding process by automating the assignment of accurate codes to diverse healthcare texts. The system is designed to process a variety of textual inputs, ranging from clinical narratives to medical reports, with the capacity to comprehend the intricacies of medical terminology, context, and coding conventions. By incorporating pre-trained word representations, MEDCODE aims to harness the power of semantic understanding, enabling the system to interpret and contextualize medical text in a manner that transcends traditional rule-based coding approaches. The web application developed as part of the MEDCODE system mirrors a user-friendly healthcare documentation platform, providing a seamless interface for healthcare professionals to input textual information and receive automated, accurate medical code suggestions. The heart of the MEDCODE model lies in its dual sub-network architecture, combining the strengths of bidirectional Long-Short Term Memory (BiLSTM) and Convolutional Neural Network (CNN) components.

The BiLSTM sub-network is tailored to capture the contextual and semantic intricacies inherent in healthcare narratives, ensuring a comprehensive understanding of the nuances embedded in the text. Simultaneously, the CNN sub-network focuses on the unique features of medical coding, extracting relevant information that pertains specifically to the assignment of codes. By emphasizing both semantic relationships and coding-specific features, the MEDCODE system transcends the limitations of conventional coding approaches, presenting a holistic solution that aligns with the complexities of healthcare documentation.

This paper lies in its contribution to advancing the field of medical coding through the integration of Natural Language Processing (NLP) techniques. Specifically, the development of the MEDCODE system offers a transformative approach to streamline and enhance medical coding processes by automating the assignment of accurate codes to diverse healthcare texts. Healthcare professionals, insurers, and regulatory bodies can benefit from this research by utilizing the MEDCODE system to improve documentation accuracy, streamline administrative tasks, and ultimately enhance patient care. Moreover, the proposed neural network architecture and methodologies presented in this paper pave the way for further research and development in the intersection of NLP and healthcare documentation, driving innovation in the field.

II. LITERATURE SURVEY

A. ICD-10 CLASSIFICATION

The paper [12] proposes an innovative method leveraging deep learning for the automatic assignment of International Classification of Diseases (ICD)-10 codes to medical records. To grasp the significance of this research, it's crucial to understand the pivotal role ICD-10 plays in healthcare. ICD-10 is a globally standardized system used for classifying diseases and injuries. It employs alphanumeric codes to represent diagnoses, procedures, and other medical conditions, ensuring consistent data collection, analysis, and reporting across

healthcare systems. The accuracy of ICD-10 coding is paramount for various purposes, including billing and reimbursement, public health surveillance, and clinical research. Manual ICD-10 coding is a cumbersome and error-prone process, posing challenges such as coding inaccuracies, delays, and high costs for healthcare institutions [6]. The paper responds to these challenges by proposing a deep learning-based solution [12]. The authors developed a model capable of automatically analyzing medical records, extracting relevant information, and assigning ICD-10 codes with high accuracy. This approach presents potential benefits, including increased accuracy due to the model's ability to comprehend complex relationships between medical terms and codes, improved efficiency by reducing manual coding efforts, and reduced costs for healthcare institutions. The research is significant as it showcases the transformative potential of deep learning in enhancing ICD-10 coding accuracy and efficiency. The proposed model yielded promising results, outperforming traditional machine learning methods. Moreover, the study emphasizes the importance of considering specific medical topics and discourse structures within medical records for precise coding. Despite the promise of deep learning-based ICD-10 coding, ethical and practical considerations must be addressed before widespread adoption. Issues such as data privacy and security, transparency, and explainability of model decisions, as well as the need for human oversight and validation, require careful attention. The paper makes a compelling case for the revolutionary impact of deep learning on ICD-10 coding, offering the potential to enhance healthcare efficiency, accuracy, and ultimately, patient care. However, responsible implementation of this technology demands a thorough consideration of ethical and practical concerns within the healthcare system.

B. MIMIC-III

The implementation of MIMIC III [5][3] involves utilizing its comprehensive dataset for healthcare research and analysis. Researchers and data scientists access the de-identified patient data to conduct studies related to critical care, patient outcomes, and healthcare analytics. The database's key features include patient information from over 40,000 ICU admissions between 2001 and 2012, encompassing a diverse range of medical conditions. MIMIC III provides a detailed set of clinical variables, including physiological measurements, laboratory results, medications, and interventions, allowing for in-depth analyses [7][11]. Its temporal data captures the progression of patient care over time, enabling researchers to examine trends, disease trajectories, and the impact of various medical interventions. Researchers often leverage MIMIC III to develop predictive models, study treatment effectiveness, and gain insights into clinical decision-making within intensive care settings. As a valuable research resource, MIMIC III contributes to advancing knowledge in critical care and supports evidence-based practices in healthcare. It is important to note that while MIMIC III offers rich clinical data, it is distinct from tools or databases used specifically in medical coding processes [4][12].

C. ELECTRONIC HEALTH RECORD

The primary focus of the paper [2] is the development of a machine learning model tailored for the automatic assignment of International Classification of Diseases (ICD) codes to Electronic Health Records (EHRs). The success of such a model is intricately tied to the diversity and comprehensiveness of the EHR data it processes. Clinical notes stand out as the mainstay of EHR data, encompassing a range of documents like discharge summaries, physician notes, progress notes, and consultation reports. These clinical notes serve as extensive repositories capturing crucial details of a patient's medical history, including diagnoses, procedures, medications, and treatment plans. Complementing the wealth of information contained in clinical notes, the study may also

draw on additional data points within EHRs. This could include diverse elements such as laboratory results, imaging reports, and medication orders, depending on the chosen model architecture and the specific preprocessing steps applied to the EHR data. Within the context of ICD classification, the paper likely incorporates several key components from the EHR content [2][6]. Existing ICD codes embedded within clinical notes represent the formal diagnoses and procedures documented by healthcare professionals. Moreover, the model may leverage natural language processing techniques to extract relevant medical terms and phrases from textual content, offering insights into diagnoses and procedures not explicitly coded with ICDs. The temporal dimension of EHR data is also emphasized, involving the analysis of the chronological sequence of events in a patient's medical history, the timing of diagnoses and procedures, and the evolution of symptoms over time. The content of EHRs assumes a pivotal role in enhancing the accuracy and effectiveness of the multi-label ICD classification model. Detailed clinical notes provide the model with a holistic understanding of a patient's health, facilitating the precise identification of relevant diagnoses and procedures. The incorporation of temporal context enriches the model's comprehension of the dynamic nature of medical conditions and their evolution over time. However, it is imperative to acknowledge potential limitations associated with EHR data. The quality and completeness of clinical notes may vary, impacting the model's performance if information is inaccurate or incomplete. Coding practices among healthcare professionals might introduce inconsistencies in ICD assignment, influencing the model's training and accuracy. Privacy concerns arise with the sharing and analysis of EHR data, demanding strict adherence to ethical guidelines to safeguard patient privacy. In navigating these intricacies, understanding the specific content and limitations of EHRs becomes essential for researchers developing robust machine learning models. This understanding ensures the ethical and responsible use of EHR data, contributing to advancements in medical coding practices and, consequently, improved healthcare outcomes and decision-making.

D. ICD-11 CLASSIFICATION

The International Classification of Diseases, 11th revision (ICD-11), stands as a crucial framework for standardizing the coding of diseases and health-related issues globally. Despite its significance, the manual assignment of ICD-11 codes can be a laborious task susceptible to errors. To address this challenge, the paper[9] introduces an innovative approach that leverages WordNet, a comprehensive lexical database, and the inherent structure of ICD-11 itself. The cornerstone of this approach is CodeNet, a network designed to bridge the gap between textual descriptions and ICD-11 codes. CodeNet capitalizes on the hierarchical structure of ICD-11, encompassing chapters, categories, and codes, to establish a structured foundation. Simultaneously, it integrates WordNet semantics, which encapsulates relationships between words, including synonyms, hypernyms, and hyponyms. By intertwining these resources, CodeNet establishes meaningful connections between medical entities in ICD-11 and their semantic equivalents in WordNet. This allows the network to comprehend the nuances of language in medical texts and accurately map them to the appropriate ICD-11 codes. The paper introduces an algorithmic approach that utilizes CodeNet to extract potential ICD-11 codes from textual input. The algorithm involves entity recognition, identifying medical entities within the text, WordNet mapping to connect these entities with their conceptual counterparts, CodeNet traversal based on mapped concepts, and candidate generation by identifying the most relevant ICD-11 codes. This approach transcends conventional keyword matching, incorporating rich semantic relationships between medical terms for more accurate and nuanced code suggestions. The implications of these findings for healthcare data management are profound.

Firstly, the proposed system significantly reduces coding errors by suggesting relevant codes based on semantic understanding, thereby enhancing the reliability and accuracy of health data. Secondly, the automation of code suggestion streamlines the coding process, saving valuable time for healthcare professionals and improving overall efficiency. Lastly, more accurate and consistent coding facilitates better analysis of healthcare trends, enabling informed decision-making and resource allocation. In conclusion, the paper's exploration of ICD-11 contents within the context of computer-assisted coding using WordNet semantics represents a promising advancement in healthcare data management. By intelligently bridging the gap between textual descriptions and standardized codes, this approach has the potential to elevate data quality, enhance efficiency, and ultimately contribute to improved patient care.

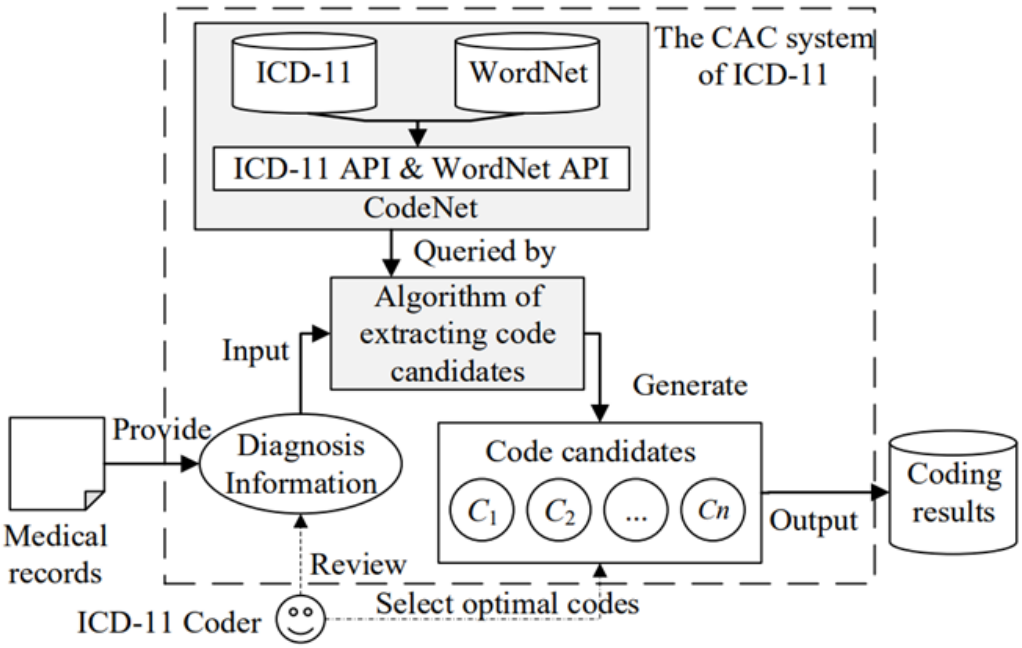


Fig. 1: ICD-11 CLASSIFICATION [9]

III. PROPOSED SYSTEM

The proposed system, MEDCODE (Medical Coding Neural Network), represents a groundbreaking approach to streamline and enhance medical coding processes in the ever-evolving healthcare landscape. Leveraging state-of-the-art Natural Language Processing (NLP) techniques, MEDCODE aims to automate the accurate assignment of standardized codes to diverse healthcare texts, ranging from electronic health records to clinical notes and medical reports. At its core, MEDCODE features a dual sub-network architecture comprising Bidirectional Long-Short Term Memory (BiLSTM) and Convolutional Neural Network (CNN) components. The BiLSTM sub-network is adept at capturing the nuanced contextual and semantic intricacies inherent in healthcare narratives, ensuring a comprehensive understanding of the text. Simultaneously, the CNN sub-network focuses on extracting coding-specific features, emphasizing information crucial for precise code assignment. By integrating these components, MEDCODE transcends the limitations of traditional coding approaches, presenting a holistic solution tailored to the complexities of healthcare documentation. This innovative system also incorporates pre-trained word representations to facilitate semantic understanding, enabling it to interpret and contextualize medical text in a manner that surpasses rule-based coding methods.

Moreover, MEDCODE offers a user-friendly web interface mirroring a healthcare documentation platform, providing healthcare professionals with a seamless experience for inputting textual information and receiving automated, accurate medical code suggestions. With its comprehensive approach to coding, MEDCODE holds the promise of significantly enhancing efficiency and accuracy in medical coding processes, ultimately contributing to improved patient care and outcomes within the healthcare ecosystem.

A. Architecture

The MEDCODE (Medical Coding Neural Network) is an innovative system that revolutionizes medical coding through advanced NLP techniques. Its dual sub-network architecture, combining BiLSTM and CNN components, enables comprehensive understanding of healthcare narratives and extraction of coding-specific features. The convolutional layers and pooling operations are tailored to capture both sentence and word embedding dimensions, enhancing feature extraction. The proposed VCPCNN architecture aims to overcome the limitations of traditional methods, leveraging deep learning advancements for comprehensive sentiment analysis in medical coding. The network's flexibility allows it to adapt to varying sentence structures and languages, making it a promising approach for nuanced sentiment classification in healthcare texts. The web application interface provides healthcare professionals with user-friendly access to input text and receive accurate code suggestions, streamlining coding processes and improving documentation accuracy. MEDCODE represents a transformative approach to medical coding, poised to enhance efficiency in healthcare documentation.

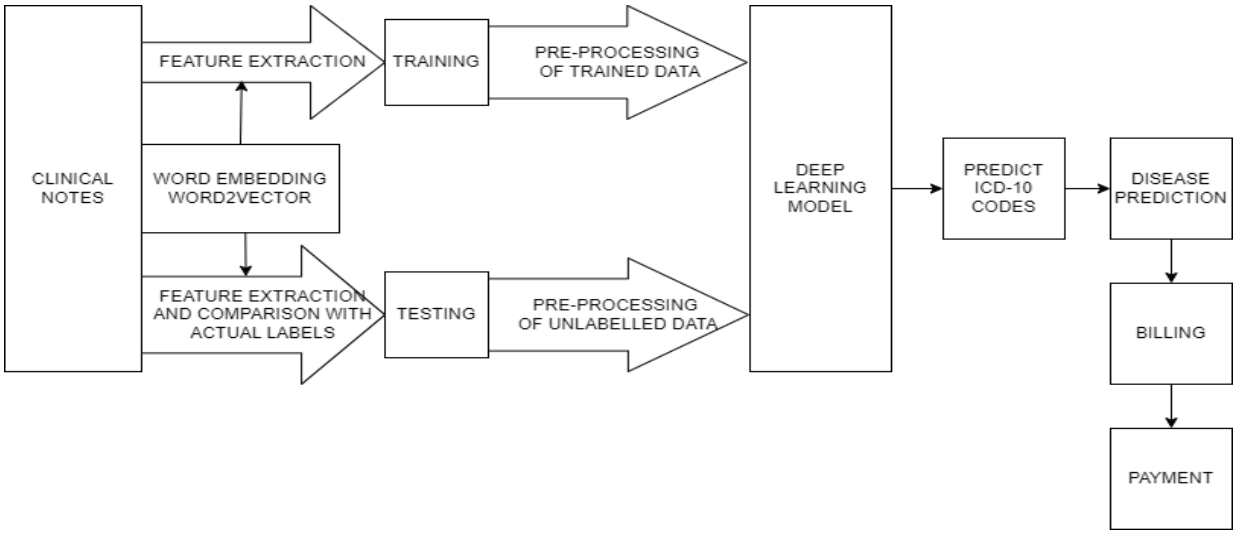


Fig-1: Architecture (MEDCODE)

B. Implementation

The data collection process for ICD-10 involves aggregating information from various sources. These encompass healthcare facilities like hospitals and clinics, where patient diagnoses and procedures are documented using ICD-10 codes. Additionally, electronic health record (EHR) systems and health information exchanges (HIEs) store comprehensive patient data, often coded with ICD-10 classifications. National health surveys, insurance claims data, and research databases also provide rich sources of ICD-10-coded information, offering insights into disease prevalence, healthcare utilization, and research trends. Public health reporting by agencies and international organizations further contributes to global health monitoring and policymaking efforts. By

accessing and consolidating data from these diverse sources, researchers and policymakers gain a holistic understanding of disease burden, treatment patterns, and public health trends, facilitating evidence-based decision-making in healthcare Preprocessing of ICD-10 coding is essential to ensure the accuracy, consistency, and usability of healthcare data. This process involves several steps, including data cleaning to remove inconsistencies and errors, normalization to standardize code formats, and validation to verify code accuracy according to coding guidelines. Understanding the hierarchical structure of ICD-10 codes allows for effective organization and grouping, simplifying analysis and navigation. Mapping ICD-10 codes to other coding systems enhances interoperability and integration with different healthcare data sources. Feature engineering techniques can extract valuable attributes from ICD-10 codes, providing insights for analysis and predictive modeling. By conducting thorough preprocessing, healthcare professionals and researchers can leverage ICD-10 coded data more effectively for clinical decision support, epidemiological studies, and healthcare analytics. Feature extraction from ICD-10 codes process involves various approaches such as calculating the frequency of each code to understand disease prevalence, determining hierarchical levels to assess diagnostic granularity, and grouping related codes to simplify analysis. By extracting these features, healthcare professionals and researchers can better understand disease patterns, healthcare utilization, and patient outcomes, ultimately leading to improved patient care and population health management.

Training a Random Forest classifier for ICD coding involves several key steps to develop an accurate predictive model. Initially, the dataset containing medical records or diagnoses coded with ICD classifications is prepared by cleaning the data and handling missing values. Relevant features are then selected from the dataset, such as patient demographics and clinical variables. The dataset is split into training and testing sets, with the former used to train the Random Forest classifier. During training, the algorithm constructs multiple decision trees based on random subsets of features and samples. Hyperparameters of the classifier are tuned to optimize performance, and the model is validated using techniques like cross-validation to assess its generalization ability. Evaluation metrics such as accuracy and F1-score are used to evaluate the model's performance on the testing set. Once trained, the Random Forest classifier can be deployed in healthcare settings to automate ICD coding, enhancing efficiency and accuracy in healthcare documentation and billing processes.

TextCNN Mode utilizes a single convolution layer and one-dimensional convolution aligned with sentence length. However, limitations arise in handling semantic dimensions. VCPCNN Model helps to address TextCNN's limitations, the Variable Convolution and Pooling Convolutional Neural Network (VCPCNN) is introduced. It includes two convolution kernel sizes for convolution in the word embedding dimension. VCPCNN-1D involves convolution in both sentence length and word embedding dimensions, with two cases: DIFF convolution (no relationship between dimensions) and SAME convolution (relationship between adjacent dimensions). VCPCNN-2D adds a convolution operation with a kernel size of $d \times 1$.

Automating the billing process in medical coding involves seamlessly integrating electronic health record systems with billing software. As medical coders assign diagnosis and procedural codes, the system automatically maps them to corresponding billing codes and calculates costs. Goods and Services Tax (GST) is applied where applicable. The system then generates itemized bills detailing diagnoses, medications, procedures, costs, and total amounts due. Bills are securely delivered to patients via email or patient portals. This automation ensures accuracy, compliance, and efficiency in billing, enhancing the overall patient experience and facilitating timely payments for healthcare services.

REFERENCE	TITLE	TECHNIQUE	MERITS	DEMERITS
[1]	Automatic medical code assignment via deep learning approach for intelligent healthcare	• Medical topic mining • Cross-textual attention mechanism • Auxiliary coding	• More explainable • More applicable to multiple languages.	• Data Dependency • Black box nature
[2]	ICD-10 Coding of Spanish Electronic Discharge Summaries: An Extreme Classification Problem	• ElMo embeddings • Attention mechanisms • Data augmentation • Ensemble methods	• Improve accuracy and efficiency of ICD coding • Outperforms existing methods	• Difficult to interpret predictions. • Only evaluated on a single dataset
[3]	UMLS mapping and Word embeddings for ICD code assignment using the MIMIC-III intensive care database	• Combination of Support Vector Machines (SVM) • FastText with Unified Medical Language System (UMLS) metathesaurus mappings	• Efficient ICD coding, • Accurate • Better diagnosis and treatment.	• Developed and evaluated for the English language only
[4]	Leveraging Semantics in WordNet to Facilitate the Computer-Assisted Coding of ICD-11	• Improved Mask RCNN • Res-Net • SSD	• More applicable to multiple languages. • More explainable	• Less accurate • Less efficient • More complex to implement

Table1:Comparative Study

IV. CONCLUSION

In conclusion, the "Medical Coding Analyzer" harnessing the capabilities of the Random Forest algorithm proves to be an indispensable tool for discerning and categorizing emotions within the medical coding realm, specifically tailored for utilization with the C-10 dataset. The incorporation of the emotion detection feature within the system holds tremendous potential for end users. It facilitates a nuanced understanding of the emotions conveyed in their comments and discussions related to medical coding. This functionality is especially pertinent for professionals in medical coding, allowing them to gauge sentiment, address concerns, and optimize their communication strategies. The analytics feature embedded in the application empowers users to delve into the emotional responses associated with medical coding content. By providing the means to analyze

and interpret sentiments expressed in comments, users can make informed decisions regarding the retention or removal of specific content. This feature contributes significantly to fostering a positive and well-informed online presence within the medical coding community. In summary, the "Medical Coding Analyzer" project underscores the effectiveness of the Random Forest algorithm in decoding emotions within the unique context of medical coding discussions. The web application emerges as an essential resource for individuals and organizations engaged in medical coding, offering valuable insights into the emotional nuances of their content and interactions.

V. REFERENCES

[1] Fei Teng, Zheng Ma, Jie Chen, Ming Xiao, and Lufei Huang, “Automatic medical code assignment via deep learning approach for intelligent healthcare”, Auckland University of Technology, May 26,2020 from IEEE xplore

[2] Blanco Almagro, M., Unanue, R. M., Fresno, V., & Montalvo, S. (2020). "ICD-10 Coding of Spanish Electronic Discharge Summaries: An Extreme Classification Problem." IEEE Access, 8, 100073–100083.

[3] Henning Schafer , and Christoph M. Friedrich, “UMLS mapping and Word embeddings for ICD code assignment using the MIMIC-III intensive care database”, ©2019 IEEE.

[4] Donghua Chen, Runtong Zhang, “Leveraging Semantics in WordNet to Facilitate the Computer-Assisted Coding of ICD-11”, IEEE Journal of Biomedical and Health Informatics, 2019.

[5] Muslim, A., Mutiara, A. B., Suhendra, A., & Oswari, T. (2018). "Expert Mapping Development System with Disease Searching Symptom Based on ICD 10." 2018 Third International Conference on Informatics and Computing (ICIC).

[6] Chen, J., Teng, F., Ma, Z., Chen, L., Huang, L., & Li, X. (2019). "A Multi-channel Convolutional Neural Network for ICD Coding." IEEE.

[7] Siangchin, N., & Samanchuen, T. (2019). "Chatbot Implementation for ICD-10 Recommendation System." IEEE.

[8] Ahmed, A. S., & Saifuddin, K. M. (2017). "A Simplistic, Effective, and Adaptive Approach towards Classifying Medical Records according to ICD-10 using Machine Learning for Efficient Statistics." IEEE.

[9] Almagro, M., Unanue, R. M., Fresno, V., & Montalvo, S. (2020). "ICD-10 Coding of Spanish Electronic Discharge Summaries: An Extreme Classification Problem." IEEE Access, 8, 100073–100083

[10] Hsu, C., Chang, P., & Chang, A. (2020). "Multi-Label Classification of ICD Coding Using Deep Learning." IEEE.

[11] Teng, F., Ma, Z., Chen, J., Xiao, M., & Huang, L. (Year). "Automatic Medical Code Assignment via Deep Learning Approach for Intelligent Healthcare." 2020 IEEE.

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

Enhancing EV Charging With Smart Prediction

Mr Amel Austine*[1], Abhijith Rajesh[2], Ian Antony[2], Joseph Peter[2], Ruben Manoj[2]
*[1]Assistant Professor, Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Vazhakulam, Kerala, India
[2]Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Vazhakulam, Kerala, India

ABSTRACT
The project revolutionizes the electric vehicle charging experience by seamlessly integrating real time range monitoring and personalized user preferences. Upon user registration, individuals input their electric vehicle details and specify their expected range. The system estimates and continuously monitors the vehicle's range, triggering a low-range alert when necessary. Users express preferences, such as being movie enthusiasts,shopping enthusiasts etc.. The project then predicts and reserves a charging station based on the user's real-time location, expected range, and preferences, with a unique emphasis on proximity to cinema theaters for movie lovers. The reserved station is communicated to the user, who receives navigation assistance to the location. During the charging process, users have access to entertainment options related to nearby cinema theaters. The system notifies users upon completion of the charging process, ensuring a personalized and enjoyable electric vehicle charging experience

Keywords: Electric Vehicle, Electric Vehicle Charging Station, Machine Learning

I. INTRODUCTION

In today's transportation sector, the shift towards electric vehicles (EVs) signifies a pivotal move towards environmental sustainability by tackling pressing ecological issues. Yet, transitioning to electric mobility comes with significant obstacles, chiefly among them being the development of a dependable and effective charging infrastructure. This project sets out to explore the integration of advanced technology with EV charging systems, with the goal of overcoming the obstacles hindering the adoption of electric mobility.

Central to achieving this vision is the innovation in predictive charging systems, an endeavor that necessitates a deep dive into a spectrum of computational techniques, from sophisticated machine learning models to advanced predictive analytics. These technologies stand at the core of our approach, designed to preemptively address the concerns of battery range anxiety and charging station availability. By harnessing these computational methods, the project not only aims to enhance the precision of charging need forecasts but also to tailor the charging infrastructure to the nuanced demands of EV users.

By weaving together the threads of advanced technology, user experience, and real-time adaptability, this project aspires to redefine the narrative around electric mobility. It aims to present a comprehensive solution that not only mitigates the current challenges faced by EV users but also paves the way for a future where electric vehicles are an integral, unencumbered part of our sustainable transportation ecosystem. Through this endeavor, the project positions itself as a cornerstone in the ongoing dialogue on leveraging technology to foster a more sustainable, efficient, and user-friendly future for electric mobility.

The document is organized as follows to systematically present the Electric Vehicle (EV) Charging Enhancement System project: Section II introduces the methodologies employed, including Geographic Information Systems (GIS), machine learning algorithms for Predictive Charging Demand Forecasting, and User Preference Analysis, outlining their roles in enhancing the EV charging experience. Section III presents a comparative analysis of these methodologies, assessing their effectiveness, limitations, and applicability in optimizing the charging process, personalizing user experiences, and addressing real-world constraints. This analysis is aimed at providing valuable insights for both academics and industry professionals, facilitating the selection of appropriate technologies and strategies for specific EV charging scenarios. It positions the section as a crucial tool for informed decision-making in the evolving field of electric mobility. Section IV introduces the proposed system, whereas section V provides the paper's conclusion by summarizing the findings and outcomes from the research and development of the Enhanced EV Charging System.

II. LITERATURE SURVEY

A. Machine Learning-Based Method for Remaining Range Prediction of Electric Vehicles

In exploration of mitigating electric vehicle (EV) range anxiety, a significant obstacle to their widespread adoption, we introduce a novel approach through the development of a blended machine learning model, as detailed in [1]. This model, leveraging historical driving data, adeptly merges the strengths of Extreme Gradient Boosting Regression Tree (XGBoost) and Light Gradient Boosting Regression Tree (LightGBM) to predict the remaining driving range of EVs with unparalleled accuracy. The intricacies of driving distance, influenced by factors such as cumulative output energy of the motor and battery, varied driving patterns, and battery temperature, are captured and analyzed within this framework. A cornerstone of our approach is the innovative "anchor (baseline) based" strategy, designed to rectify the skewed distribution of datasets, thereby ensuring the robustness and reliability of the model's predictions. The experimental validation of this model, as presented in [1], confirms its exceptional performance, notably reducing the prediction error range to [-0.8, 0.8] and significantly outperforming existing predictive models. This breakthrough not only exemplifies the efficacy of combining XGBoost and LightGBM in an anchor-based model but also illuminates the path towards alleviating range anxiety among EV users, potentially accelerating the adoption of EVs by enhancing user trust in their driving range capabilities[1].

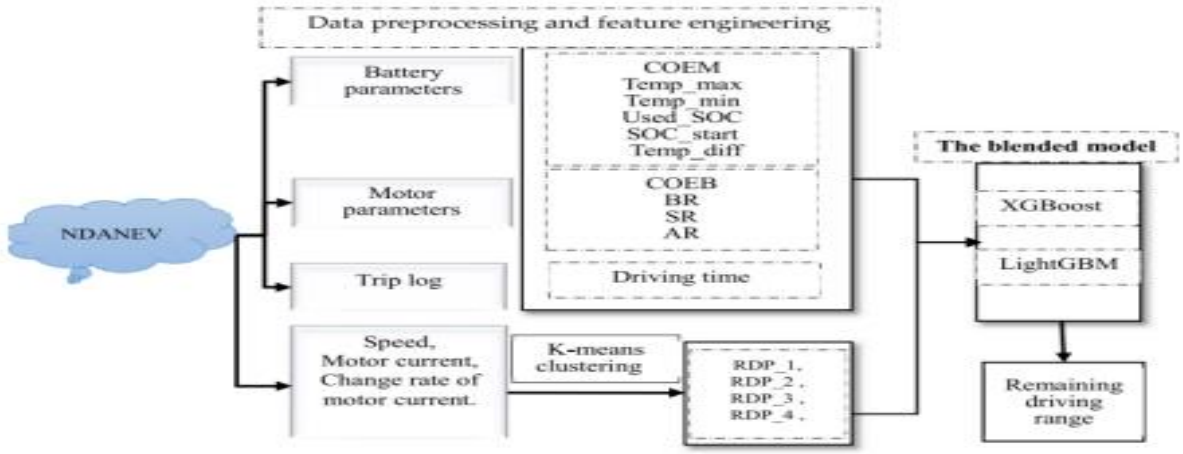


Fig. 1 Architecture diagram

B. Prediction of Electric Vehicle Charging Stations Distribution Using Machine Learning

In the evolving landscape of transportation, electric vehicles (EVs) are increasingly becoming the preferred choice among consumers, highlighting the urgent need for the expansion of electric vehicle charging stations (EVCS). Addressing this need involves the strategic placement of new charging stations, a challenge that has been tackled through the creation of a predictive model, as detailed in [2]. This model employs a dataset designed to forecast optimal locations for new EVCS installations, utilizing machine learning algorithms such as K-nearest neighbors, logistic regression, neural networks, and support vector machines. The city of Dubai, UAE, serves as the focal point for this case study, with the model considering variables like population density, points of interest (POI), and the presence of security cameras to recommend suitable locations for EVCS.

The methodology encompasses the generation of a dataset from geographic coordinates of 162 existing EVCS and the utilization of ArcGIS software to map population density across various Dubai clusters, marking existing EVCS locations with red indicators. The significance of potential locations is assessed based on these criteria, including the strategic presence of security cameras, to ensure safety and convenience. This approach led to the creation of a balanced dataset, divided into training, validation, and testing sets, with a notable validation accuracy of 89% achieved using the K-Nearest Neighbors model.

This predictive framework not only facilitates the identification of prime locations for EVCS in densely populated and frequently visited areas but also integrates safety considerations through the inclusion of security camera presence. By leveraging machine learning classification models, the study outlined in [2] provides a comprehensive analysis and solution to the challenge of expanding EVCS infrastructure in urban areas like Dubai, underscoring the importance of data-driven decisionmaking in enhancing EV charging facilities and supporting the transition towards sustainable mobility solutions.

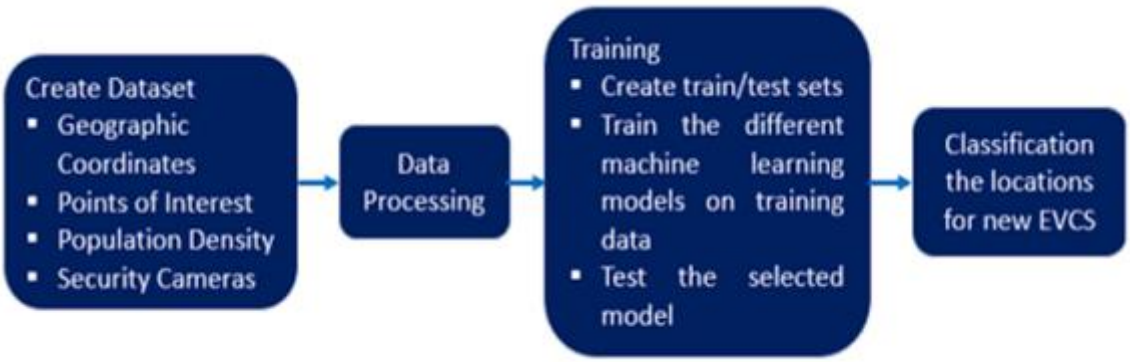


Fig. 2: Framework of proposed methodology

C. Prediction of EV Energy Consumption Using Random Forest and XGBoost

In response to the growing adoption of electric vehicles (EVs) as a sustainable substitute for fossil fueldependent transportation, the challenge of efficiently managing EV charging station infrastructures has become increasingly prominent. While EVs are pivotal in reducing carbon emissions and diminishing the reliance on fossil fuels, the deployment and upkeep of EV charging stations necessitate significant financial and infrastructural investment. A crucial component in optimizing this infrastructure is the precise prediction of EV user behavior, such as the timing of arrivals and departures at charging stations and the specific energy requirements of each vehicle. The current landscape lacks in-depth, data-driven models capable of accurately forecasting these behaviors. Addressing this gap, our proposed model employs Artificial Neural Network (ANN) and Machine Learning (ML) models to predict EV user behaviors effectively, utilizing extensive data that includes individual instances of charging and weather conditions. The objective is to cultivate intelligent algorithms that enable the scheduling of EV charging to minimize the need for an extensive charging infrastructure, thereby fostering a more efficient and sustainable EV ecosystem. To achieve this, our model leverages the ACN-Dataset, a comprehensive dataset maintained by the California Institute of Technology, which provides real-time data on workplace charging sessions, particularly focusing on data from the Jet Propulsion Laboratory (JPL) from November 2018 to November 2019 for model training and validation. This approach is enriched by integrating weather data from the National Oceanic and Atmospheric Administration (NOAA) and calendar data, including public holidays and weekly patterns, to gain a nuanced understanding of user behavior. This dataset, encompassing data for 352 users, forms a robust foundation for our study. By incorporating these diverse data sources, our research aims to significantly enhance the prediction accuracy of user behavior, paving the way for optimized and efficient charging station management. This holistic methodology is poised to contribute substantially to reducing CO2 emissions and the dependency on fossil fuels in the transportation sector, as detailed in our conference paper [3].

D. Real-time optimization in electric vehicle stations using artificial neural networks

The study at hand introduces an innovative smart decision-making algorithm tailored for electric vehicle (EV) charging stations, aiming to minimize the total charging time by accurately predicting queuing delays. At the heart of this approach lies an Artificial Neural Network (ANN) model, which is fed with a specially curated dataset to forecast the queuing delay under conditions where the number of vehicles surpasses the available charging points. The effectiveness of this model is substantiated by an impressive accuracy rate of 89%. Further

validation is provided through comparison with a meta-heuristic optimizer, which evidences a reduction in total charging time by 2.5%, and a significant 23.9% reduction when contrasted with a basic model devoid of optimization strategies. A pivotal aspect of this research is the physical realization of the ANN model, achieved by simulating a vehicle as a transmitting node and the charging station as a receiving node, thereby demonstrating the practical applicability of the model in real-world settings. The methodology underpinning this research incorporates the generation of a dataset through the Monte Carlo Simulation Technique, encompassing variables such as distance, charge status and battery charging capability. The ANN model, optimized using the the Levenberg–Marquardt method for its efficiency, is meticulously designed to predict the minimum anticipated delay for future EVs in the queue. This prediction takes into account various scenarios, including the presence of single and multiple charging stations, aiming to significantly reduce the total charging time by factoring in both queuing delay and travel time. The innovative approach extends to the stimulation of a physical model through an embedded system prototype, which simulates the dynamics of multiple EVs and charging points. This step bridges the theoretical and practical realms, showcasing the model's capability to enhance the management of EV charging stations, as detailed in [4]

III. PROPOSED SYSTEM

A. Architecture Diagram

This innovative system marks a significant evolution in electric vehicle (EV) charging strategies, seamlessly blending machine learning's predictive prowess with an emphasis on individual customization. Engineered to forecast EV battery depletion levels, especially noticeable during weekends when driving habits may vary from weekday routines, the system utilizes cutting-edge machine learning algorithms to determine when an EV needs recharging and selects the ideal charging station based on geographical location and user preference

The EV Charging Optimization System spearheads the evolution of electric mobility by harnessing cutting-edge technologies to refine the EV charging process. Utilizing Geographic Information Systems (GIS) in tandem with advanced machine learning techniques, including Forecasting for Charging Demand and Analysis of User Preferences, this platform skillfully predicts EV charging needs and identifies the ideal locations for charging stations, considering both user preferences and geographic information.

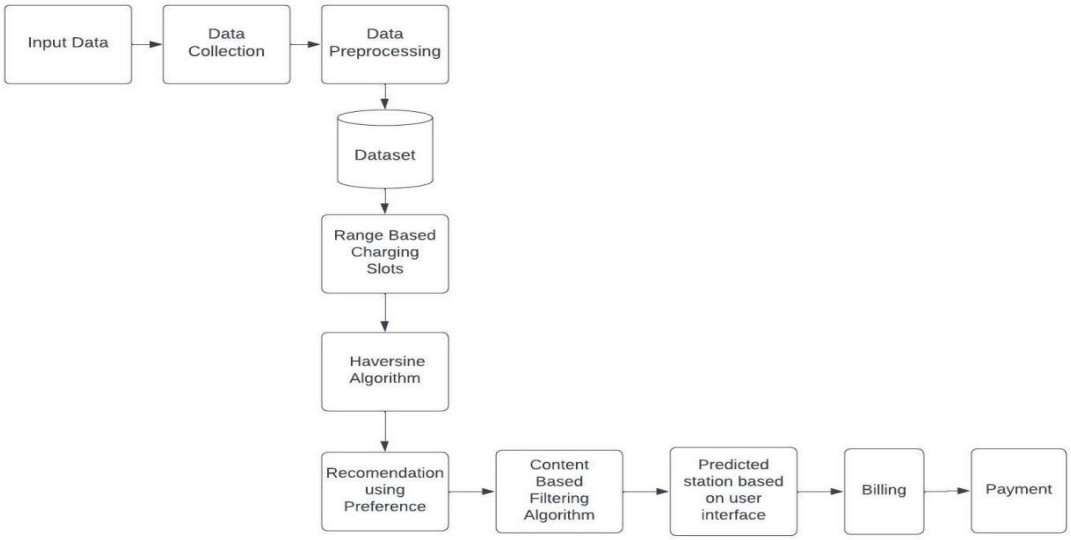


Fig. 5: Proposed System Architecture

The uniqueness of this system stems from its capability to incorporate personal preferences into the charging decision making process. For users with a penchant for leisure activities, such as visiting cinemas or dining out, it smartly prioritizes charging stations in close proximity to entertainment venues or gourmet eateries. This approach not only guarantees timely recharging of EVs but also significantly enhances the overall user experience by aligning with individual lifestyle choices.

This ensures that station recommendations are tailored to both personal desires and practical realities, optimizing both efficiency and convenience for the user. The system's user interface is designed to be both intuitive and engaging, simplifying the process of setting preferences and identifying recommended charging stations, and can be integrated into existing EV frameworks or offered as a standalone application In essence, this system elevates beyond a mere charging solution to become a lifestyle enhancement tool for EV users. By merging machine learning with a design focused on the user, it seeks to transform the EV charging experience into a seamless, enjoyable, and personalized journey

B. Implementation Details

- Charging Station Recommendation Engine: The Charging Station Suggestion System plays a crucial role in our customized electric vehicle charging solution. Utilizing an advanced mix of machine learning techniques and data centered around the user, this system recommends charging stations that best match individual preferences and actual conditions. Below is an in-depth explanation of its functionality.
- Data Collection and Analysis: Collect data on individual user preferences, lifestyle choices (like affinity for movies or dining), and typical weekend travel patterns enhancing understanding of the classroom environment triggered to notify users of any concerning trends, promoting timely medical attention.
- User Interface Development: The creation of the user interface (UI) for our tailored electric vehicle charging platform is designed to offer an interactive, straightforward, and accessible experience. This UI acts as a conduit between the advanced backend algorithms and the end-users, facilitating the entry of user preferences, the delivery of tailored charging station suggestions, and smooth interaction with the platform.
- User Interface Development: The creation of the user interface (UI) for our tailored electric vehicle charging platform is designed to offer an interactive, straightforward, and accessible experience. This UI acts as a conduit between the advanced backend algorithms and the end-users, facilitating the entry of user preferences, the delivery of tailored charging station suggestions, and smooth interaction with the platform.

IV. COMPARATIVE STUDY

Based on the reviewing of the papers, a thorough comparison analysis was conducted. This study attempted to identify and evaluate the different benefits and drawbacks that are present in every technology that was discussed in the literature.

Method	Advantage	Disadvantage
[1]Machine Learning-Based Method for Remaining Range Prediction of Electric Vehicles	•Provides a scalable and efficient approach to EV charging station allocation •Reduced range anxiety	• Real-world uncertainties • Computational complexity
[2]Prediction of Electric Vehicle Charging Stations Distribution Using Machine Learning	•Data-driven decision making •Improved user experience	• Model limitation • Data dependency
[3]Prediction of EV Energy Consumption Using Random Forest and XGBoost	•High accuracy •Robustness	• Complexity • Data requireme-nts
[4]Real-time optimization in electric vehicle stations using artificial neural networks	•High accuracy in predicting queuing delay of EV •Outperforms other optimization approaches in terms of queuing delay reduction	•Still in early stages of development, so further validation in real-world EV stations is needed

Table 1 : Comparative Study

V. CONCLUSION

The Intelligent Predictive Charging initiative marks a pivotal development in the progression of electric vehicle (EV) technology, blending machine learning, design centered on the user, and the enhancement of efficiencyThe Intelligent Predictive Charging initiative marks a pivotal development in the progression of electric vehicle (EV) technology, blending machine learning, design centered on the user, and the enhancement of efficiency.

These innovations guarantee a charging experience that is both reliable and efficient, tailored to the unique needs and preferences of each user. The integration of precise location tracking, along with an advanced booking system and proactive notifications for users, further refines the charging experience by reducing waiting times and dispelling uncertainties. This initiative extends its reach beyond merely combining technology with user experience, delving into the broader domain of sustainable transportation. By rendering EVs more accessible and centered on the user, it plays a part in fostering a cleaner and more sustainable future, in line with global efforts to combat climate change. The effective rollout of the Intelligent Predictive Charging framework highlights the significant role of innovative technologies in transforming the EV ecosystem.

Looking forward, there is immense opportunity for continued innovation and improvement. Future projects could explore the integration of renewable energy sources into the charging network, enhancing the environmental sustainability of the system. Additionally, as EV technology evolves, the flexibility and scalability of the framework will be crucial in accommodating new advancements and meeting changing user

needs. This initiative not only provides a robust solution to present challenges but also lays the foundation for ongoing advancements in electric mobility. The contributions of this project are anticipated to have a lasting impact on the community, encouraging further innovation towards a more sustainable and efficient transportation system.

VI. REFERENCES

[1] Liang Zhao, Wei Yao, Yu Wang and Jie Hu "Machine LearningBased Method for Remaining Range Prediction of Electric Vehicles" in Elsevier, November 2020 . doi:10.1109/ACCESS.2020.3039815.

[2] Marwa Chendeb El Rai, Sabina Abdul Hadi, Haitham Abu Damis and Amjad Gawanmeh. "Prediction of Electric Vehicle Charging Stations Distribution Using Machine Learning" in 2022 5th International Conference on Signal Processing and Information Security (ICSPIS), December 2022. doi:10.1109/ICSPIS57063.2022.10002556.

[3] Harshit Rathore, Hemant Kumar Meena and Prerna Jain. "Prediction of EV Energy Consumption Using Random Forest and XGBoost" in 2023 International Conference on Power Electronics and Energy (ICPEE), January 2023. doi:10.1109/ICPEE54198.2023.10060798.

[4] Elkasrawy, M.A., Abdellatif, S.O., Ebrahim, G.A. et al. Real-time optimization in electric vehicle stations using artificial neural networks (2023). doi:10.1007/s00202-022-01647-9.

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

Prediction of Health Using Wearable Devices with Machine Learning Techniques

Ms Ierin Babu*[1], Aashuthosh S[2], Anandhu S[2], Cristin Siljo[2], George Geo[2]
*[1]Assistant Professor, Department of CSE, VJCET Ernakulum, Kerala, India
[2]Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Ernakulam, Kerala, India

ABSTRACT
The Health Monitoring System addresses the limitations of traditional checkups amidst busy schedules. The system leverages various features like step tracking, heart rate monitoring, oxygen level monitoring, activity tracking, and calorie count with food composition analysis. Using MLP(Multi Layer Perceptron) , the app predicts potential health risks based on user data and recommends personalized exercises and dietary plans to prevent them. This user-friendly and data-driven tool empowers individuals to take control of their health and prioritize it even in hectic lives, propelling us towards a healthier future.

Keywords: Machine learning, Deep learning, Recurrent Neural Network, Gated Recurrent Unit, Long Short-Term Memory.

I. INTRODUCTION

The convergence of wearable technology and Machine Learning(ML) holds immense potential for revolutionizing healthcare. Wearables, with their ability to continuously track various physiological and behavioural metrics, provide a rich source of data that can be harnessed to predict and manage health conditions. ML, on the other hand, offers powerful algorithms capable of identifying patterns and relationships within this data, leading to personalized insights and predictions about a user's health.

The given system delves into the exciting realm of predicting health using wearable devices and machine learning techniques. We will explore the various types of wearable sensors and the data they collect, delve into the capabilities of machine learning algorithms in analysing this data, and examine the potential applications of this emerging technology in various healthcare domains. We will also discuss the challenges and limitations associated with this approach, and outline future directions for research and development.

As we embark on this journey, it is important to remember that the integration of wearables and machine learning for health prediction has the potential to democratize healthcare, empowering individuals to take proactive control of their health and well-being. By harnessing the power of these technologies, we can pave the way for a future of personalized, preventative, and predictive healthcare, ultimately leading to better health outcomes for all.

II. LITERATURE SURVEY

A. Medical History Based Disease Prediction

The methodology implemented in [1] as depicted in Figure 1 employs a hybrid approach through the seamless integration of three distinct models: NAIS, DeepICF, and DeepFM, for disease prediction. In the initial phase of data processing, diseases are represented effectively through dense vectors utilizing an embedding matrix. The extraction of relations is facilitated by Multi-Layer Perceptron (MLP) incorporating batch normalization for enhanced performance. The integration of low and high-order relations is achieved through the utilization of a factorization machine. The final step involves determining the likelihood of disease occurrence through the application of a sigmoid function. To optimize prediction accuracy, the loss function uses a L2 regularization term. This comprehensive methodology aims to leverage the strengths of each individual model in a synergistic manner, enhancing the overall predictive capabilities for disease outcomes. The hybrid methodology adopted in this recommendation system amalgamates various recommendation techniques, ensuring a well-rounded and nuanced array of suggestions. DeepICF, one of the integrated models, capitalizes on both item relations and user-item representations to elevate the precision and relevance of recommendations. The inclusion of Factorization Machine (FM) plays a crucial role in managing high-dimensional and sparse features, contributing to a more robust and precise modeling of the recommendation landscape. In the realm of features, the system takes into consideration a spectrum of user attributes and preferences. Specifically, it incorporates correlations such as gender and purchase preferences to deliver more personalized and tailored recommendations. This incorporation of feature insights introduces a layer of complexity to the recommendation process, addressing individual user characteristics and preferences. The factorization process within the model employs auxiliary vectors to tackle feature sparsity challenges. This approach proves effective in capturing essential feature relations, ensuring a more comprehensive understanding of intricate patterns within the data. By utilizing auxiliary vectors, the model enhances its ability to address sparse features, contributing to the overall effectiveness of the recommendation system. In essence, the amalgamation of these components results in a recommendation system that not only integrates diverse methods but also considers critical feature insights and employs sophisticated factorization techniques for a more refined and accurate user experience.

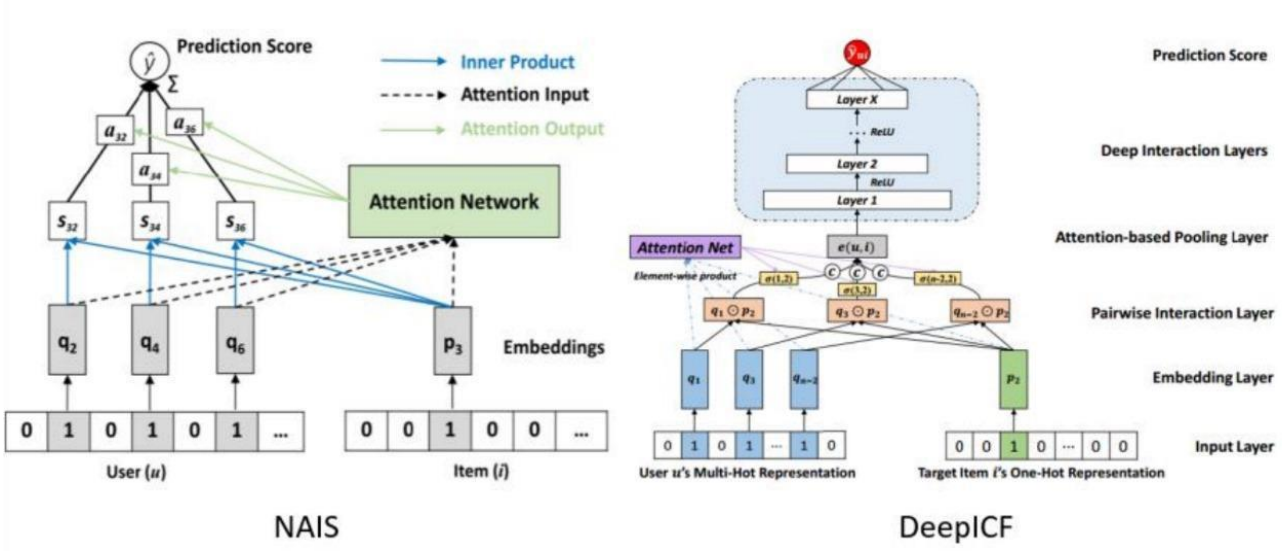


Fig. 1 NAIS and DEEPICF Framework

B. IoMT Assisted Framework

The recommendation system utilizes a hybrid strategy that combines various recommendation methods to deliver a comprehensive array of suggestions. One integrated model, DeepICF, enhances recommendations by leveraging item relations and user-item representations, leading to improved precision and relevance in the suggested content. Additionally, FM is employed to effectively handle high-dimensional and sparse features, thereby enhancing the overall modeling of the recommendation landscape. In the feature domain, the system takes into consideration an assortment of user attributes and preferences. Specifically, the incorporation of correlations such as gender and purchase preferences introduces complexity to the recommendation process, resulting in more customized suggestions for users. Furthermore, the factorization process incorporates the use of auxiliary vectors designed to address feature sparsity challenges.

This strategic use of auxiliary vectors assists in capturing crucial feature relations, contributing to a more nuanced understanding of intricate patterns within the data. The system discussed in [2] as depicted in Figure 2 is developed to fulfill the crucial requirement for personalized dietary recommendations, emphasizing the related health advantages. Employing advanced machine learning and deep learning techniques, the project aims to offer customized guidance rooted in individual health conditions and preferences. Both user information and product details are taken into account for short and long-term scenarios, prioritizing the protection of patients with diseases. To address individual needs, the system incorporates a registry based alerting system through an AI-driven automatic notification system. A refined approach to Long Short-Term Memory (LSTM) enhances precision in meeting patient needs, while the integration of contextual and social information contributes to improved accuracy in the input data received. Significantly, the system broadens its coverage to encompass a wider range of health conditions compared to earlier systems.

The data processing methodology embraces a structured approach that integrates normalization and encoding techniques to establish uniformity, ensuring adaptability across a spectrum of machine learning (ML) algorithms. The identification of crucial features within the dataset is achieved using the Random Forest Classifier, providing insights into feature importance. A variety of deep learning classifiers, including the MLP, Recurrent Neural Network (RNN), LSTM, and Gated Recurrent Unit (GRU), are utilized to harness the capabilities of neural networks for classification tasks. Concurrently, traditional machine learning classifiers like Naive Bayes and Logistic Regression are implemented for effective classification, considering factors such as feature independence and probability calculations.

Data collection for this study involved gathering information from a sample of 30 patients using IoT and Cloud Methods. Subsequent data processing procedures encompassed normalization for scaling, encoding to convert nominal values into numeric ones, and Optimal Feature Visualization to identify crucial features within the dataset. In the phase of data classification, Deep Learning methods such as MLP , RNN , LSTM, and GRU were employed to classify the data effectively.

Moving forward, the study involves a comparison between Machine Learning and Deep Learning Models, with a recommendation of the best performing model determined through the application of evaluation metrics across all suggested models. By combining the strengths of FM and Deep Neural Networks (DNN), the proposed model excels in handling both high-order relations among disease features and the low-order combinations leading to specific diseases. The model's comprehensiveness is enhanced through experiments on real-world datasets predicting potential phenotypes based on observed gene phenotypes. Notable improvements are attributed to the incorporation of an attention network that weights historical diseases, reducing noise from

irrelevant diseases. Additionally, the joint learning of deep and FM parts merges low and high-order relations, yielding superior predictive performance. While accurate disease prediction is crucial for medical examination assistance

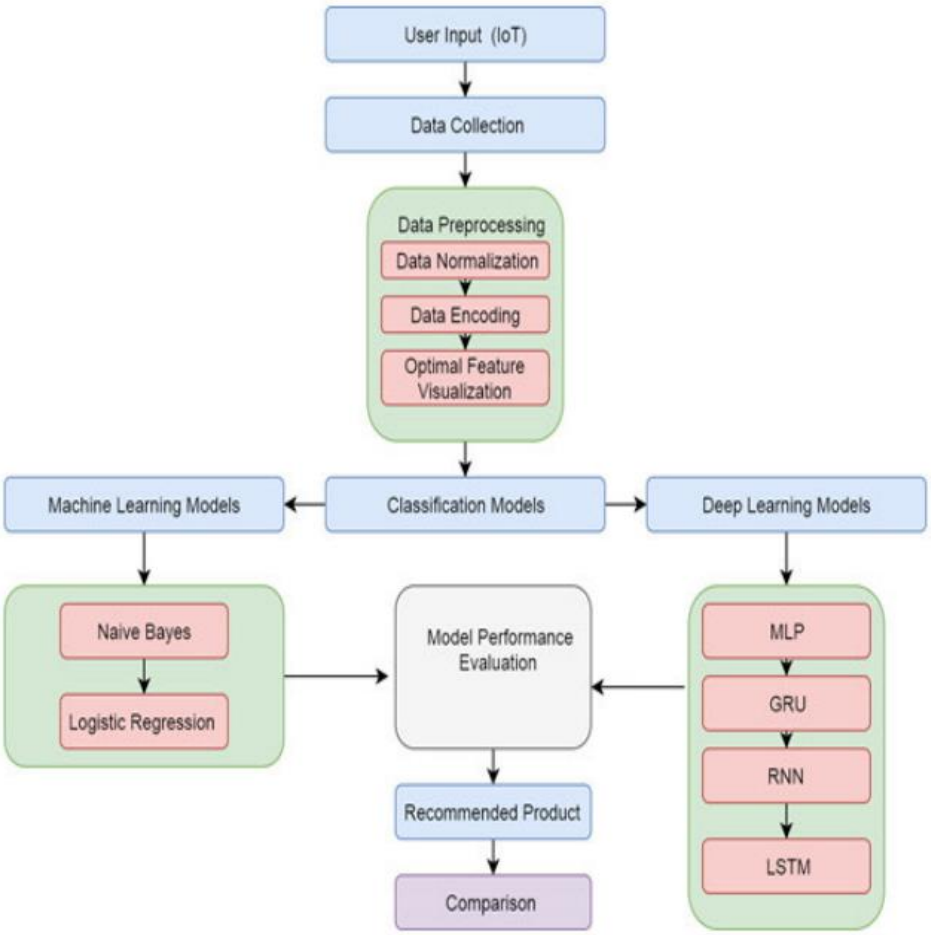


Fig. 2: Workflow Model of IoMT Assisted Framework

C. Novel Time-Aware Food Recommender System

This system as depicted in [3] as depicted in Figure 3 proposes a Novel Time-aware food recommender system based on Deep Learning and Graph Clustering (TDLGC) to address the shortcomings of previous food recommender systems. TDLGC tackles four issues: 1) It considers both user and food similarity based on ingredients. 2) It incorporates a time-aware similarity measure to handle changing preferences. 3) It utilizes a trust network to address cold start problems. 4) It leverages user communities for recommendations.

Suggesting food items tailored to a user's preferences and past interactions involves a sophisticated approach that integrates content analysis and deep learning techniques. By examining the ingredients and nutritional information of various food items, a graph clustering method is employed to identify clusters of similar items and users. Leveraging a deep-learning model, the relationships between food items and users are systematically learned. These learned relationships are then harnessed to provide personalized recommendations, ensuring that users receive suggestions that align with their tastes and dietary preferences. This comprehensive system

combines analytical methodologies and cutting-edge technology to enhance the precision and relevance of food recommendations for individual users.

components: user-based and food content-based recommendation systems. On the left side, user-based recommendation involves computing user similarity, clustering similar users, and predicting ratings for new food items based on their cluster and ratings from others in that cluster. Simultaneously, the right side focuses on food content-based recommendation, utilizing deep embeddings to capture food item characteristics, calculate item similarity, cluster similar items, and predict ratings for users based on their food item cluster and ratings from similar users. In the center, the Trust Relationship Network signifies social connections, and the Graph Representation of Users illustrates users as nodes interconnected by trust-based edges. The integration of user-based and food content-based recommendations, along with trust relationships, empowers the Top-N Recommendation block to suggest personalized food items. The system's adaptability is emphasized by the optional and combinable nature of user-based and food content-based recommendations. Although not explicitly depicted, the system accommodates temporal factors in crafting personalized recommendations, with arrows denoting information flow and highlighting the dynamic interplay within this comprehensive recommendation framework.

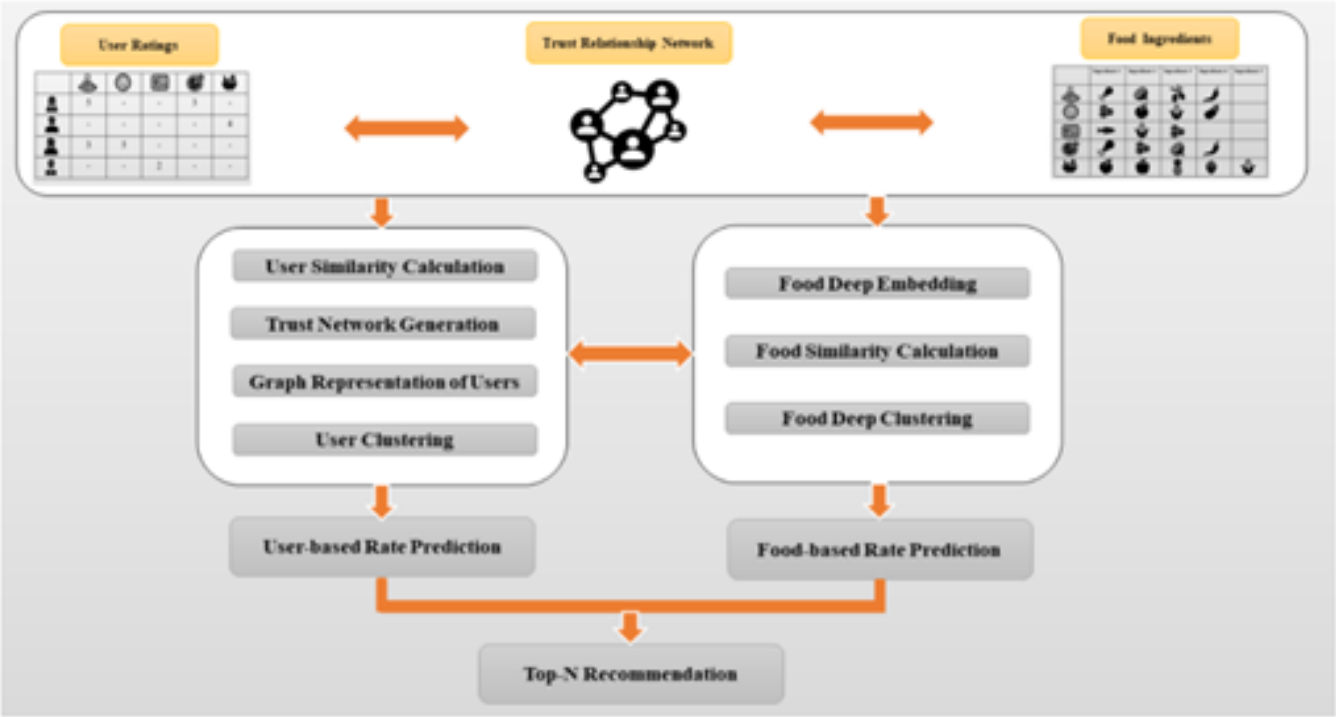


Fig 3 Framework of Food Recommender System

D. LSTM-RNN Based Multi Disease Prediction Framework

Chronic diseases often strike silently, leaving individuals and healthcare systems unprepared. This paper proposes a novel approach to predict the risk of multiple diseases simultaneously, leveraging the power of LSTM, Recurrent Neural Networks and longitudinal medical records. Unlike static snapshots, the LSTM architecture implemented in [4] "remembers" medical history, capturing the dynamic interplay of factors over time. This enables multifaceted predictions and time-aware insights, leading to more accurate and timely risk assessments. Evaluated on real-world datasets, the model demonstrates superior performance, paving the way

for personalized medicine and potentially life-saving interventions. This research opens doors to a future where prevention triumphs over late-stage treatment, empowering individuals to take control of their health journey

Predicting multiple diseases before symptoms whisper their arrival. This research dives into this frontier, leveraging the wealth of longitudinal medical records. Gone are static snapshots; instead, the dynamic interplay of factors across time takes centre stage. Methods implemented such as LSTM,RNN, AI masters at understanding sequential data. Imagine remembering past medical events to diagnose current illnesses; LSTMs do the same, analysing your unique medical journey like a seasoned doctor.

But before this analysis, data prep is crucial. Age and other continuous variables get normalized to speak the same language. Categorical variables like gender are one-hot encoded, transforming distinct categories into digestible formats. Now, the data is ready for the LSTM units, the model's core. These units excel at capturing long-term dependencies

The final stage involves translating the model's insights into probabilities. A sigmoid activation function converts the model's calculations into a range from 0 (no risk) to 1 (high risk). By entering the binary cross-entropy loss function, meticulously penalizing the model for deviations between its predictions and the actual disease labels. The smaller the penalty, the closer the model gets to predicting risks accurately and timely.

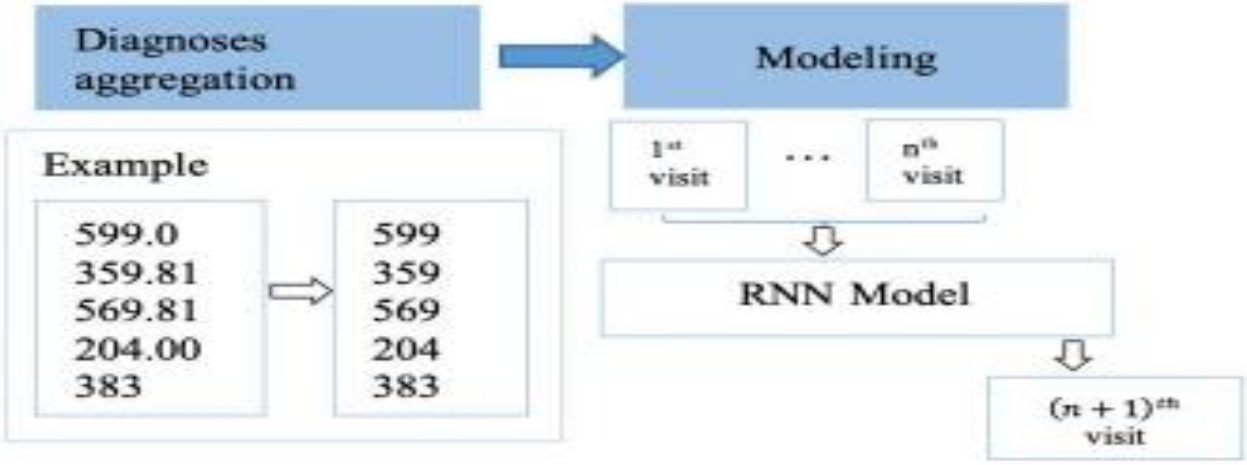


Fig 4 Framework for multi-disease risk predictions

III. PROPOSED SYSTEM

A. Architecture Diagram

The proposed framework introduces a health monitoring system centred on a wearable device, seamlessly integrating diverse health data for risk assessment and personalized recommendations. Comprising a smartwatch, mobile app, and a centralized database, the system captures the user's heart rate, spO2, and temperature. The wearable device transmits this data to the mobile app for processing, where comprehensive health reports are generated. The app not only delivers feedback to users but also suggests dietary adjustments and lifestyle modifications to enhance overall well-being. Crucially, the centralized database securely stores all health-related information, allowing healthcare providers access for more informed care provision. The anticipated benefits encompass improved health data tracking, personalized health feedback, tailored lifestyle recommendations, and streamlined health data sharing with healthcare professionals.

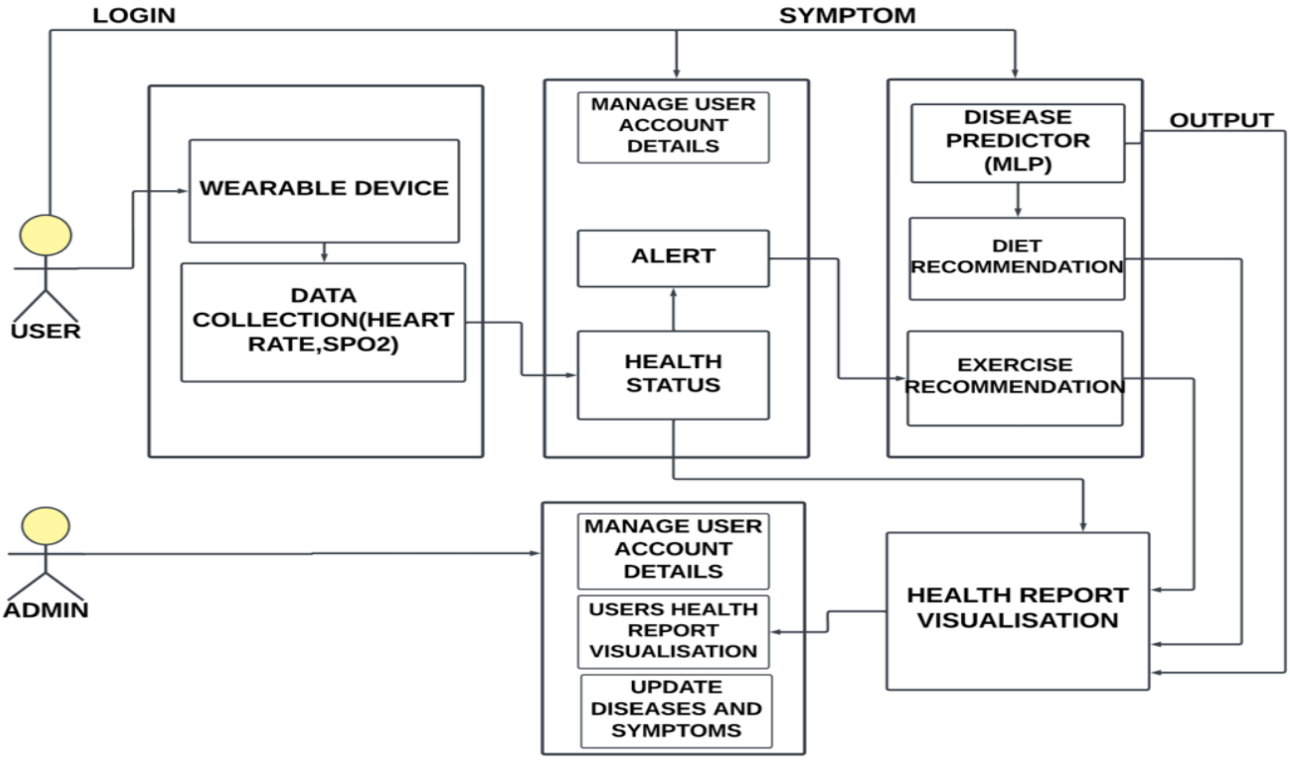


Fig. 5: Proposed System Architecture

The proposed system as depicted in Fig 5 takes a proactive approach to wellness by utilizing wearable devices to continuously monitor user health. Users can access the system through a secure login portal where they can manage their account details and even input any symptoms they might be experiencing. The wearable device acts as a vital partner, constantly collecting health data like heart rate and blood oxygen saturation. This data is then securely transferred to the system where it gets analysed by a powerful disease prediction system. It also employs MLP(MultiLayer Perceptron) for training and testing evaluation for disease prediction, and scans the data to identify potential health risks.

Based on the analysis, the system provides users with valuable health insights. This can include a clear picture of their overall health status, along with alerts for any concerning data points that might require attention. To further empower users, the system can even generate personalized recommendations. This could include suggesting dietary plans tailored to their health needs or recommending exercise routines that can help them improve their well-being. Users can continue to manage their account details within this section, ensuring they stay in control of their health information.

On the administrative side, the system offers a dedicated area for admins to view insightful visualizations of user health reports. This allows them to monitor the overall health trends of the user base. Additionally, admins can update the system's database of diseases and symptoms. This ensures the system stays current with the latest medical knowledge and can provide the most accurate predictions and recommendations possible. Overall, this health system creates a collaborative environment where wearable devices, user input, and advanced AI analysis work together to empower users to take charge of their health and well-being.

B. Implementation Details

Fig. 6: Data Acquisition from Sensors MAX30102 and ESP32 microcontroller

The proposed system explores a novel health system designed to empower individuals through proactive health monitoring. The system leverages wearable sensor technology, like the MAX30102 sensor, to capture vital health metrics such as blood oxygen saturation (SpO2), heart rate and temperature. This data is then wirelessly transmitted via Wi-Fi to a central application for further analysis.

It employs a multi-step approach to translate raw sensor data into actionable insights. First, the collected data undergoes a meticulous evaluation process. This evaluation compares the data against established health benchmarks, enabling proactive health monitoring and potential intervention strategies if necessary. Alerts can be triggered to notify users of any concerning trends, promoting timely medical attention.

Next, the data undergoes preprocessing to ensure its accuracy and suitability for further analysis. Techniques such as null value checking and removal, as well as duplicate data identification and elimination, are employed to refine the dataset. This meticulous cleaning process guarantees the integrity of the data used for subsequent applications.

Finally, the system leverages machine learning for deeper insights. Categorical data, such as health evaluation results, is transformed into numerical formats for compatibility with machine learning algorithms. This process, known as training conversion, allows the system to identify patterns and trends within the data. Balance of the dataset is crucial for performance and accuracy , if dataset is imbalanced then oversampling is applied to resolve the issue. Random Oversampling is one such method These insights can then be used to formulate personalized recommendations and predictions regarding future health needs.

Overall, this health system fosters a proactive approach to healthcare. By harnessing the power of wearable technology, wireless data transmission, and machine learning, the system empowers individuals to take charge of their well-being and make informed decisions about their health journey.

IV. COMPARATIVE STUDY

Based on the reviewing of the papers, a thorough comparison analysis was conducted. This study attempted to identify and evaluate the different benefits and drawbacks that are present in every technology that was discussed in the literature. Table 1, which methodically provides a comprehensive summary of the observed contrasts and captures the complex subtleties of the technologies covered in the corresponding papers, is essential to this analysis. After a methodical examination and evaluation, the table provides significant value as a tool for comprehending the technologies under investigation and provides information about their individual advantages and disadvantages.

Paper	Models	Advantage	Disadvantage
A Medical-History Based Potential Disease Prediction Algorithm	DeepICF, FM	Enhances suggestion accuracy. Factorization handles feature sparsity and improves robustness of the system.	Complexity due to multiple models being integrated.
Realizing an Efficient IoMT-Assisted Patient Diet Recommendation System Through Machine Learning Mode	MLP, LSTM, GRU, RNN	Early Detection of Health Issues Can automatically learn intricate data models.	Effective training requires large data. Complex Implementation
A Novel Time-Aware Food Recommender-System Based on Deep Learning and Graph Clustering	Graph Clustering, Food Clustering,	Considers Time Factor and ser preferences	It requires large data. It has high computational complexity It suffers from Cold Start Problem
Long Short-Term Memory Recurrent Neural Networks for Multiple Diseases Risk Prediction by Leveraging Longitudinal Medical Records	LSTM, RNN	It handles long sequences of data. It avoids the vanishing gradient problem. It can handle variable-length sequences.	Higher data sparsity. Long Training Time. Suffers from Overfitting Hyperparameter Tuning.

Table 1 : Comparative Study

V. REFERENCES

[1] Iwendi, Celestine, et al. ”Realizing an efficient IoMT-assisted patient diet recommendation system through machine learning model.” IEEE access 8 (2020): 28462-28474

[2] Hong, Wenxing, et al. ”A medical-history-based potential disease prediction algorithm.” Ieee Access 7 (2019): 131094-131101.

[3] FARRAHI, VAHID. ”A Novel Time-Aware Food Recommender-System Based on Deep Learning and Graph Clustering.”

[4] Wang, Tingyan, Yuanxin Tian, and Robin G. Qiu. ”Long short-term memory recurrent neural networks for multiple diseases risk prediction by leveraging longitudinal medical records.” IEEE journal of biomedical and health informatics 24.8 (2019): 2337-2346

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

Shadow Removal from Document and Conversion to Digital Format

Asha Joseph[1], Amala John[2], Liya Mariya Abraham[2], Misty Sunny[2], Theresa Polychan[2]
*[1]Assistant Professor, Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Muvattupuzha, Kerala, India
[2]Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Muvattupuzha, Kerala, India

ABSTRACT
Shadows are prevalent in nature. Shadows often occur when we capture the documents using casual equipment. The Frequency-Aware Shadow Erasing Net (FSENet) for shadow elimination in document images, presents a solution to the prevalent issue of shadows which are captured using casual equipment. A specialized deep learning network can be utilized to decompose images into low and high-frequency components, and thus the system can effectively separate shadows from document content. Integration of Optical Character Recognition (OCR) software transforms documents into searchable digital formats, thus boosting accessibility and usability. The proposed system has the potential to redefine document processing by enhancing readability, improving OCR precision, and simplifying content modification, thus making a significant impact on the quality and usability of digital document copies.

keywords— Laplacian Pyramids Decomposition, Frequency- aware Shadow Erasing Network, Optical Character Recognition

I. INTRODUCTION

Shadow removal is a crucial task in computer vision, significantly enhancing image visual quality with widespread applications. Deep learning methods, particularly those based on Convolutional Neural Networks (CNNs), have emerged as powerful tools for this purpose, operating on either paired or unpaired data. In this research landscape, various innovative approaches address challenges in shadow removal, each offering unique contributions. One approach introduces a Lightness-Guided Shadow Removal Network (LG- Shadow Net) [1], emphasizing training on unpaired data. This method employs two CNN modules, with the first compensating for lightness and the second refining shadow removal guided by lightness information. The incorporation of a loss function leveraging color priors enhances overall performance, as evidenced by extensive experiments across well-known datasets. Another proposed method, the Decoupled Multi-Task Network (DMTN) [2], introduces feature decomposition to explicitly learn shadow removal, shadow matte estimation, and shadow image reconstruction.

In contrast, physics-based [3] approach focuses on shadow image decomposition, SP-Net and I-Net are utilized in tandem to accurately forecast shadow parameters, generate matte layers, and further enhance shadow-free images. By incorporating both fully supervised and weakly supervised frameworks, this model demonstrates superior performance in terms of shadow removal accuracy and preservation of intricate details, surpassing the capabilities of existing methods. However, challenges persist within these methods. Computational complexity, limitations in handling multiple light sources, and the need for fine-tuning in specific scenarios pose practical constraints. Furthermore, the black-box nature of deep learning models hinders interpretability and understanding of the decision-making process. This framework leveraging a conditional random field (CRF) model presents a promising approach in the domain of single-color image shadow detection and removal, offering innovative solutions in the field. Feature learning through ConvNets and hierarchical color transfer strategies contribute to improved accuracies in shadow detection tasks across various datasets. A novel deep reciprocal network (DRNet) [4] introduces a reciprocal learning approach, involving a shadow detector and remover. Despite notable successes, challenges arise in accurately estimating shadow masks and preserving color in certain scenarios, emphasizing the need for further refinement. A new method for removing document shadows introduces the SD7K dataset and Frequency-Aware Shadow Erasing Net (FSENet), addressing dataset limitations [5]. By utilizing a Laplacian Pyramid and multi-frequency components, this approach achieves computational efficiency, addressing challenges associated with real-world document shadow removal.

In this survey paper, section II consists of literature survey that explores various related works, section III consists of proposed system that describes the methodology and architecture in detail, section IV consists of a comparative study of the related works discussing the various advantages and disadvantages, section V consists of conclusion.

II. LITERATURE SURVEY

There are a variety of ways to remove shadows, from simple lighting adjustments and post-processing in software tools to advanced machine learning techniques to improve image quality and analytical precision. All of these methods work together to reduce the problems caused by shadows, providing a wide range of effective solutions in image processing and computer vision. Here are some methods which are found to be very helpful in removing shadows.

A. PHYSICS-BASED SHADOW IMAGE DECOMPOSITION

In [3], H. Le *et al.* describe the approach to removing shadows from images using deep learning. The method outlines the utilization of two deep neural networks, SP-Net for predicting shadow parameters and appearance, followed by the application of another network named I-Net to further refine the obtained results. The SP-Net predicts the shadow parameters, which represent the intensity and color changes caused by shadows. The Matte Network (M-Net) predicts the matte layer, which represents the soft shadows and the boundaries between shadow and non-shadow regions. The Inpainting Network (I-Net) refines the recovered shadow-free image by filling in any missing or corrupted details. Here shadow illumination model is incorporated into the well-known image decomposition system. Here two supervised frameworks are explained. In a fully supervised network, this approach utilizes SP-Net and M-Net trained on paired shadow and shadow-free images. Also training I-Net to refine shadow removal results further. Here we use the Trust Region method for

optimization. In weakly- supervised network utilizes unpaired shadow and shadow-free patches to train the shadow decomposition. In conjunction with SP-Net, an M-Net Adversarial Network (D-Net) serves as a discriminator, effectively distinguishing between generated and authentic shadow-free patches. Outperforms other methods in shadow removal with high detail preservation, handling fine-grained details and complex textures. Utilizes a physics-based model for shadow formation, enhancing resilience to diverse lighting conditions and object geometries. They have a multi-stage approach for shadow removal, allowing flexibility in handling intricate shapes and varying intensities, with added capability for temporal shadow dynamics in dynamic scenes. The computational complexity of this method, due to its deep learning architecture and multi-stage processing, poses challenges for real-time applications or resource-constrained environments. Additionally, the model's limitations, such as reliance on a single light source assumption and the need for fine-tuning based on specific scenarios, reduce its plug-and-play nature and interpretability.

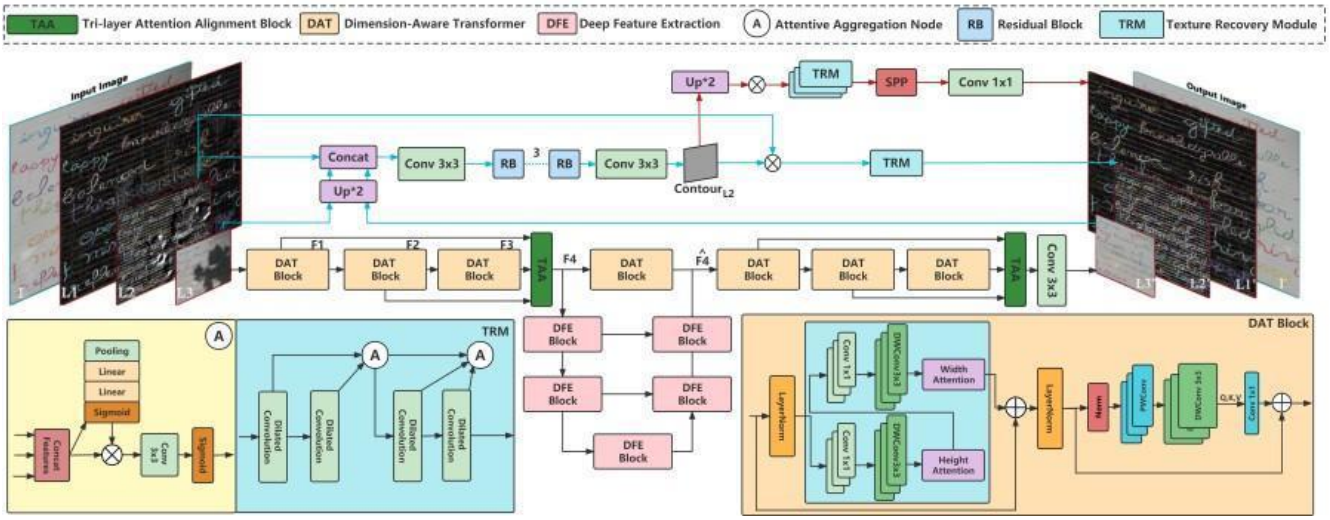
B. BAYESIAN SHADOW REMOVAL

In [6], S. H. Khan *et al.* presented a framework for precise pixel-level detection and localization of shadows in single- color images. The particular problem is formulated as a conditional distribution using a conditional random field (CRF) model. The energy function in the CRF model incorporates unary and pairwise potentials. Feature learning for unary predictions involves training two ConvNets separately for regional and boundary potentials. The ConvNets operates on patches extracted around points of interest. The CRF model is used for shadow contour generation, and the proposed shadow removal and matting framework includes identifying umbra, penumbra, and non-shadowed regions. A hierarchical color transfer strategy is employed, and a generalized shadow generation model is then introduced. The process involves Bayesian shadow removal and matting, optimizing parameters through a maximum a posteriori estimate. The resulting shadow-less image undergoes further processing for boundary enhancement using Poisson smoothing. Through experimental evaluations conducted on UCF, CMU, and UIUC datasets, the effectiveness of the proposed framework is showcased, exhibiting enhanced accuracies in shadow detection tasks compared to existing state-of-the-art methods.

C. DEEP RECIPORCAL NETWORK

In [4], Lou *et al.* proposed a unique way to remove shadows in images without using pixel-coding annotations. A deep mutual network (DRNet) is needed, which is made up of SD shadow detector and SR shadow remover. Mutual learning is made up of 3 stages: Pre-learning of SR, estimation of shadow masks, and more shadow removal training. A self-paced learning approach is used to handle noisy labels educationally. The shadow elimination process is presented as an optimization problem and the model can maximize the likelihood of the determined data when given a latent shadow mask. The model struggles as it should be able to estimate shadow masks, and finds it difficult to distinguish between shadows and dark stuff, together with black stone. The second failure case is that the version does not repair the color effectively in certain cases. This indicates a need to do some work in color matching and renovating when dealing with positive types of snapshots or content material.

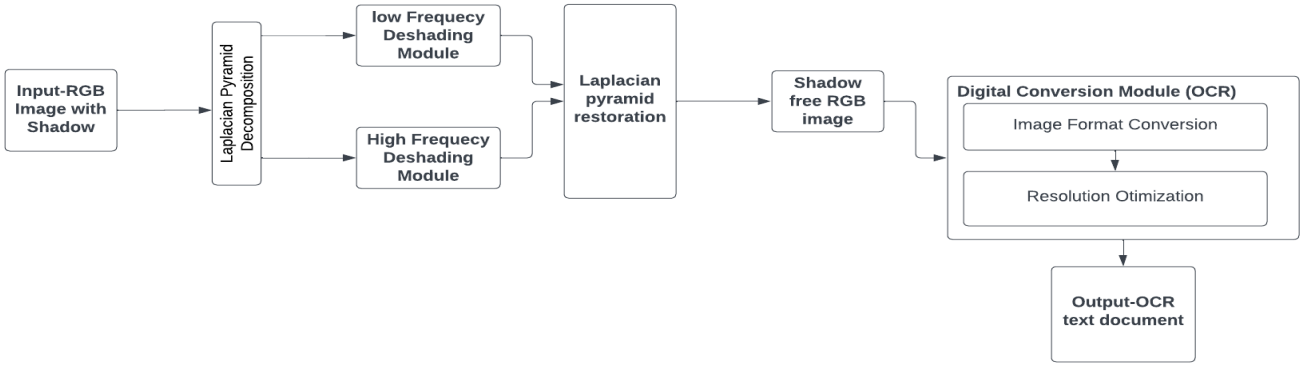
D. FREQUENCY-AWARE SHADOW ERASING NET



In [5], S. H. Khan *et al.* the recommended approach to report shadow elimination provides a large-scale, real-world dataset, SD7K, to overcome the drawbacks of current datasets, which may be small and synthesized, thus deviating from actual samples in the distribution. The Canon EOS M 6 digicam is used in controlled conditions to capture paired shadow/shadow-loose photos, taking into account lighting, digicam configurations, and more than 30 different types of occluders. The 7620 samples of datasets are used to boost the variety. The authors recommend a frequency-aware shadow erasing net (FSENet) as shown in Fig. 1., which divides the enter photograph with awesome frequency additives using the Laplacian pyramid (LP) to overcome the challenges associated with real global record shadow removal. The low-frequency find-out module deals with the low-frequency capabilities, the high-frequency recovery module deals with the high-frequency capabilities, and the new contour expansion and texture recovery module work together. The recommended approach achieves computing performance by treating each multifrequency factor one at a time and then integrating the LP additives to achieve the very last results. SmoothL1/SSIM loss features are utilized inside the education to hold picture information.

III. PROPOSED SYSTEM

The suggested method makes use of FSENet's (Frequency-Aware Shadow Erasing Net) sophisticated capabilities to solve the problem of image shadow removal. The advanced architecture of FSENet, specifically engineered for accurate feature selection, facilitates the system's ability to detect and remove shadows, hence improving the overall visual quality of the photos it processes.



A. Architecture

First, the approach uses Laplacian pyramids decomposition to divide the input photo into low-frequency and high-frequency objects. The deshading module then processes the low-frequency problem and removes the shading artifacts from the image. The high-frequency shading artifacts are eliminated by processing the excessive aspect using unique deshading modules. The two deshading module outputs are combined to form the restored picture.

B. Implementation

As shown in Fig.2., the suggested device for shadow removal in RGB images incorporates sophisticated image processing methods. To handle shadows, which sometimes appear as darker or differently colored regions inside the RGB color space, the Feature-aware Shadow Erasing Network (FSENet) is utilized. The gadget utilizes the Laplacian pyramid and frequency decomposition techniques to partition images into constituent parts primarily based on their frequency content. This enables multi-resolution illustration and captures images with remarkable details and edges.

The combined actions of the high-frequency restoration module and low-frequency deshading module, work together to reduce the influence of shadows on both finer information and global lighting. While the high-frequency restoration module selectively highlights and restores intricate details inside shadowed areas, the low-frequency deshading module adjusts normal brightness by addressing wider pixel depth differences caused by shadows. This integration makes it possible to take a photograph with less shadow and more clarity. The next stage is image reconstruction, which combines the enhanced high-frequency information with the corrected low-frequency data in an elegant manner to provide a final result that is comprehensive, logical, and aesthetically pleasing. The digital conversion module finishes the shadow removal pipeline by transforming the improved shadow-free image into a chic virtual format. This guarantees eco-friendly storage, display, and subsequent processing of the image while preserving its quality and integrity. The depiction of this all-encompassing strategy shows how well the suggested technique works to solve shadow-related issues and improve the visible clarity of RGB images.

IV. COMPARATIVE STUDY

When compared with other shadow removal methods FSENet is more efficient in removing shadows from high-resolution documents even though it is time-consuming as evident from Table 1.

Physics-based shadow image decomposition is reliant on smaller and potentially less representative datasets, which limits its ability to generalize to diverse real-world shadow scenarios.
Bayesian shadow removal is reliant on a less sophisticated method for addressing the challenges associated with real-world shadow removal, which limits its effectiveness in achieving precise and comprehensive shadow elimination.
In deep reciprocal network, there is difficulty in effectively correcting color inconsistencies, particularly in certain cases, which indicates a need for improved color matching and restoration techniques.

TABLE 1 COMPARISON OF FSENET WITH OTHER SHADOW REMOVAL MODELS

TITLE	Model	Advantages	Disadvantages
Physics-Based Shadow Image Decomposition for Shadow Removal	Physic-based image decomposition [3]	No need for shadow-free images.	Incorrect colors in the lit area.
Learning Shadow Removal from Unpaired Samples via Reciprocal Learning	DRNet [4]	Very little color distortion and very faint and dark shadows are removed to a great extent.	The estimation of a shadow mask might be wrong in the case of a dark background color.
Automatic Shadow Detection and Removal from a Single Image	Bayesian Shadow Removal [6]	Does not require any user input.	Does not perform well on curved sur-faces.

V. CONCLUSION

In conclusion, Frequency-Aware Shadow Erasing Net is used due to its advanced capabilities in handling real-world documents. FSENet's advanced design makes it easier to select features with accuracy, which improves our system's capacity to identify and eliminate shadows and enhances the overall visual quality of processed images. By using the OCR will be able to convert the image into an editable format. It can be used in various industries dealing with document digitization such as libraries, archives, and administrative offices. A comparative study was done based on different shadow removal techniques and concluded that FSENet is the best method for removing shadows from the documents.

VI. REFERENCES

[1] Z. Liu, H. Yin, Y. Mi, M. Pu, and S. Wang" Shadow Removal by a Lightness-Guided Network With Training on Unpaired Data," in IEEE Transactions on Image Processing, vol. 30, pp. 1853-1865, 2021, doi: 10.1109/TIP.2020.3048677.
[2] J. Liu, Q. Wang, H. Fan, W. Li, L. Qu and Y. Tang," A Decoupled Multi-Task Network for Shadow Removal," in IEEE Transactions on Multimedia, doi: 10.1109/TMM.2023.3252271.

[3] H. Le and D. Samaras," Physics-Based Shadow Image Decomposition for Shadow Removal," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 12, pp. 9088-9101, 1 Dec. 2022, doi: 10.1109/TPAMI.2021.3124934.

[4] W. Luo, X. Xie, K. Deng, L. Yang and J. Lai," Learning Shadow Removal From Unpaired Samples via Reciprocal Learning," in IEEE Transactions on Image Processing, vol. 32, pp. 3455-3464, 2023, doi: 10.1109/TIP.2023.3285439.

[5] Zinuo Li, Xuhang Chen, Chi-Man Pun, and Xiaodong Cun," High- Resolution Document Shadow Removal via A Large-Scale Real-World Dataset and A Frequency-Aware Shadow Erasing Net." (2023).

[6] S. H. Khan, M. Bennamoun, F. Sohel and R. Togneri," Automatic Shadow Detection and Removal from a Single Image," in IEEE Trans- actions on Pattern Analysis and Machine Intelligence, vol. 38, no. 3, pp. 431-446

[7] J. Liu, Q. Wang, H. Fan, J. Tian and Y. Tang," A Shadow Imaging Bilin- ear Model and Three-Branch Residual Network for Shadow Removal," in IEEE Transactions on Neural Networks and Learning Systems, doi: 10.1109/TNNLS.2023.3290078.

[8] M. Xu, J. Zhu, P. Lv, B. Zhou, M. F. Tappen and R. Ji," Learning- Based Shadow Recognition and Removal From Monochromatic Natural Images," in IEEE Transactions on Image Processing, vol. 26, no. 12, pp. 5811-5824, Dec. 2017

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

NFC Airport App

Ms. Ariya T. K*[1], Abel Binoy[2], Martin Antony[2], Varun Mohammed[2]
*[1]Assistant Professor, Department of Computer Science and Engineering, VJCET, Ernakulam, Kerala, India
[2]Student, Department of Computer Science Engineering, VJCET, Ernakulam, Kerala, India

ABSTRACT
The AirMate NFC AIRPORT TOOL is a mobile application that revolutionizes the airport experience for travelers through NFC technology integration. With NFC-based flight information access, electronic boarding passes, luggage tracking, and airport services, the app streamlines the travel process. Additionally, real-time flight updates, interactive airport maps, and travel assistance features enhance the traveler's experience. To further augment convenience, a centralized travel itinerary management platform is incorporated. Prioritizing security and privacy, the app employs robust measures to protect user data. The AirMate NFC AIRPORT TOOL has the potential to redefine the airport experience, providing a streamlined, secure, and personalized journey from check-in to departure.To further augment the traveler's experience, the AirMate NFC AIRPORT TOOL incorporates a centralized travel itinerary management platform. This centralized hub allows users to effortlessly manage their travel plans, including flight details, hotel reservations, rental car information, and any other pertinent travel documents, ensuring that all essential information is readily accessible in one convenient location.

Keywords: NFC, API

I. INTRODUCTION

In today's fast-paced world, travelers seek seamless and efficient airport experiences. NFC (Near Field Communication) technology is emerging as a game-changer in the aviation industry, revolutionizing the way passengers navigate airports and enhancing their overall travel experience. NFC-enabled airport apps are transforming traditional airport processes, offering a plethora of benefits to both passengers and airport operators. An NFC-enabled airport app is a mobile application that utilizes NFC technology to facilitate contactless interactions between passengers and NFC-enabled infrastructure within an airport. These interactions enable a range of functionalities, streamlining various aspects of the passenger journey, from check-in to boarding.

NFC technology relies on radio waves to transmit data between two devices when they are brought close together. NFC-enabled airport apps utilize this technology by embedding NFC chips or tags within airport infrastructure, such as kiosks, scanners, and beacons. When a passenger's NFC-enabled smartphone or passport

approaches these NFC-enabled elements, data is exchanged wirelessly, enabling seamless interactions and providing valuable information.

II. RELATED WORK

A. NFC for Smart Operations

NFC technology has been explored for streamlining various airport operational processes and workflows in recent research studies. One study designed a NFC-enabled quality inspection system that allows airport staff to log safety violations and non-conformities by tapping NFC tags on assets like ground support equipment, vehicles, and aircraft. This contactless approach improves record traceability and access.Another research work proposed a secure NFC check-in kiosk system to prevent passenger identity fraud during self-service airline check-ins. Passengers verify their identity by tapping NFC-enabled travel documents against the kiosk before printing boarding passes, adding a layer of biometric authentication.

In the area of staff access control, a pilot project implemented NFC badges for ground handling personnel. Employees tap their NFC badges on readers at airport entry points, enabling real-time tracking of their locations and efficient schedule management based on this data.Researchers have also investigated using NFC forcontactless data exchange between aviation systems. One study demonstrated secure NFC-based updates of aircraft navigation databases and flight plan loading, replacing traditional physical media transfers.

Some work has focused on using NFC for enabling smart airport maintenance workflows. Concepts like NFC tags on critical equipment allow maintenance staff to easily log service records and view updated maintenance histories via mobile devices.

B. NFC for Airport Resource Tracking

Several studies have investigated using NFC technology for tracking and managing airport resources like ground support equipment, baggage trolleys, and other mobile assets. By attaching NFC tags to these assets, their real-time locations can be monitored through an infrastructure of NFC readers installed across the airport premises. This enables optimizing asset utilization and preventing losses.One research project implemented a NFC-based baggage cart tracking system, allowing ground crew to check-in/check-out carts through NFC touch points. The system provided analytics on cart dwell times and predicted demand surges for proactive deployment.

Another study focused on using NFC tags and readers for Real-Time Location Systems (RTLS) to track aircraft mobility, ground vehicles, and fuel/cargo loading equipment. This location intelligence can streamline airport logistics and surface movement management.

C. NFC for Passenger Token/Digital Identity Management

In the paradigm of mobile digital identities and tokenization, NFC has been studied as a secure component for passenger token acquisition, verification and access control within airports. Researchers have proposed architectures where passengers can acquire verifiable digital travel credentials on their NFC-enabled mobile devices through tap interactions at airport kiosks or checkpoints.

These digital tokens can then be used for contactless verification and deriving context-aware access privileges at subsequent checkpoints using NFC handshakes with server-side identity providers. Such frameworks have been

evaluated for processes like vehicle access, lounge entry, duty-free purchases using the secured digital identity token.

D. Integration of NFC with Other Sensor Modalities

While NFC provides an intuitive interaction technology, some research has focused on fusing it with other sensor inputs from mobile devices and IoT infrastructure to create sophisticated digital environment interactions within airport terminals.

For instance, one approach combined NFC tags with Bluetooth beacons and ambient light sensors to provide multi-modal feedback to users based on their real-time context like location, ambient lighting conditions, etc. This allowed rendering adaptive audio-visual-haptic navigation assistance.Other work studied sensor fusion pipelines that utilized NFC as a proximity trigger, which then activated additional sensors like cameras for gesture tracking or using the inertial measurement unit (IMU) to gauge user mobility patterns after an NFC tag interaction.

E. NFC Enabled Smart Airports and Internet of Things (IoT)

Looking ahead, NFC is seen as an enabling technology for interconnecting intelligent sensor networks and IoT devices within smart airport environments of the future. Researchers are studying architectures where NFC can act as a last inch data link and intuitive configuration interface for airport IoT deployments.For example, NFC-based commissioning of IoT infrastructure has been conceptualized, where airport operators can enroll newly installed IoT sensor nodes into operational systems by simply tapping NFC devices against them. NFC then securely transfers provisioning metadata to the sensors.

Overall, as seen from this literature review, NFC provides a rich design surface for creating innovative digital aviation solutions by synergistically combining it with other technologies like IoT, computer vision, secure identity management and sensor fusion.

F. NFC for Disabled Passenger Assistance

Leveraging NFC's intuitive tap paradigm, researchers have explored solutions to enhance accessibility and provide seamless assistance for disabled passengers at airports. One system allowed passengers to register their special needs and accessibility requirements by tapping NFC cards during check-in. This data was shared securely across airport touchpoints to provision wheelchairs, adjust instructions and adapt self-service interfaces based on individual needs.

Other concepts include NFC-based indoor navigation solutions tailored for visually impaired passengers. These utilize patterns of strategically positioned NFC tags as tactile paths across terminals, with audio instructions triggered on smart assistants through NFC taps.

G. NFC for Secure Crew Operations

The aviation industry has strict security protocols for crew members, aircraft handling and maintenance operations. In this context, NFC technologies have been investigated as a secure, contactless mechanism for access control and data acquisition by crew members.

Certain studies proposed using NFC for electronic signature capture and secure documentation handover during pre/post-flight procedures like moving flight papers, safety checklists between aircraft crew and ground operators. NFC ensures data integrity through encrypted transmissions.

NFC badges have also been conceptualized as a secure crew identity verification mechanism that gates access to sensitive areas like aircraft galleys/cockpits based on custom permission policies configured by airlines.

III. PROPOSEDSYSTEM

A. Architecture

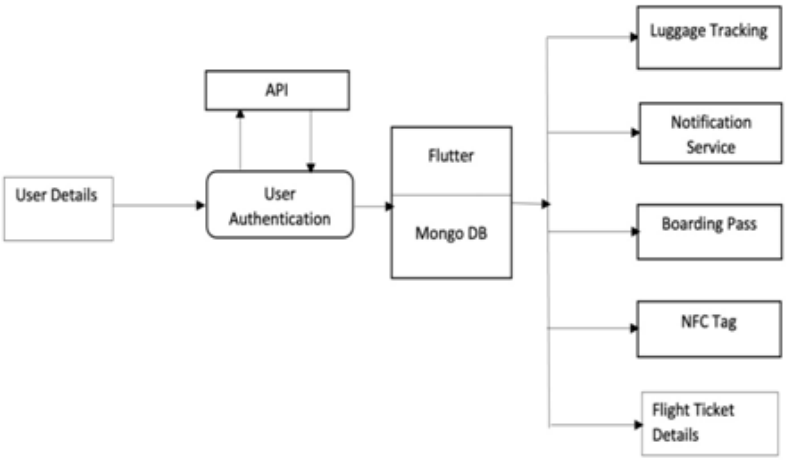


Fig.5.Architecturediagram

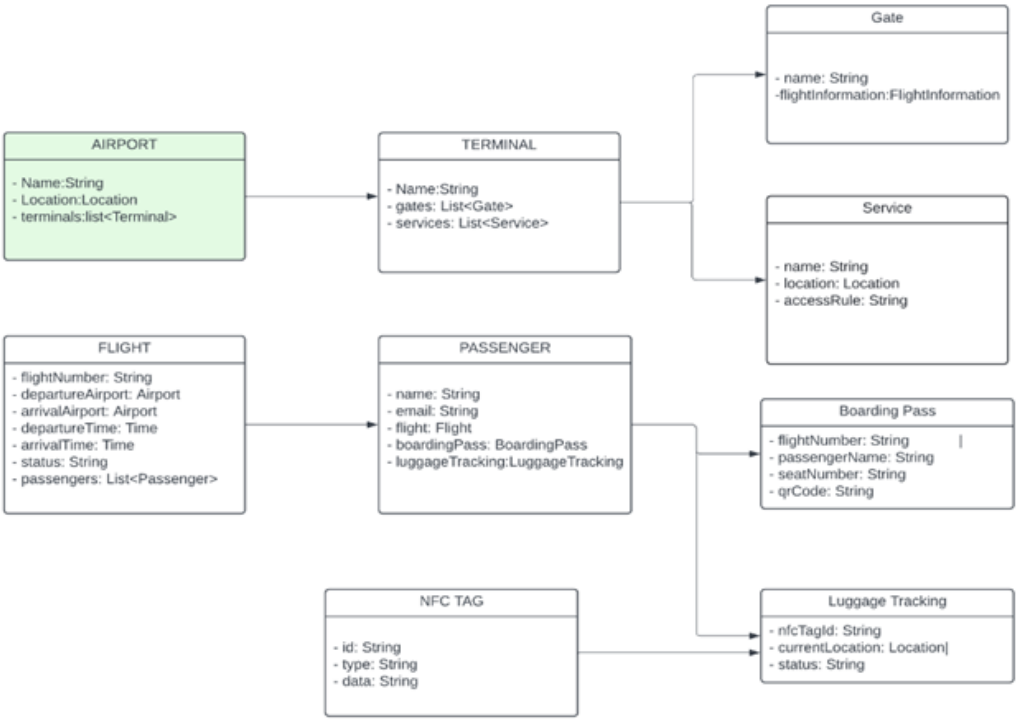


Fig6.Proposedsystemoverview

B. Implementation

Passenger Management:This includes user registration, login, and profile management features which allows users to link their flight information and manage their travel itinerary within the app.

Flight Information:This model Integrates with an API or flight data provider to retrieve real-time flight information for departure/arrival times, gate assignments, and status updates.which will display the information clearly and conveniently for users.

Boarding Pass Management:This includes Considering of integrating QR code generation and scanning capabilities to scan the boarding pass for fetching flight information.

Luggage Tracking:This model includes Developing of features for associating NFC tags with specific luggage which utilizes NFC reader libraries to scan tags and track their location within the airport (considering integrating with airport's luggage tracking system if available)and alsoprovides real-time updates on luggage status within the app.

IV. COMPARATIVESTUDY

Paper	Merit	Demerit
An ISO/IEC 7816-4 Application Layer Approach to Mitigate Relay Attacks on Near Field Communication	Effective Protection against relay attacks	Not a Complete Solution
RONFC: A Novel Enabler-Independent NFC Protocol for Mobile Transactions	Scalable to large no of users	Complexity is there for this protocol
SCSEC-Enchancing Security In Near Field Communication Through Screen Camera Communication	Simplicity	Reliance on Screen(Too much dependence on Screen)
A Novel NFC-Based Secure Protocol for Merchant Transactions	Efficient	Initial Investment

V. CONCLUSION

In conclusion, The proposed NFC AIRPORT TOOL APP has the potential to revolutionize the airport experience by providing a streamlined, secure, and personalized journey from check-in to departure. Throughthe integration of NFC technology, real-time information services, and comprehensive travel

management tools, this tool aims to address the key pain points of air travel and enhance the overall travel experience. By providing passengers with NFC-based functionalities like accessing flight information, using electronic boarding passes, tracking luggage, and seamlessly accessing airport services, the app empowers users with convenient access to essential information and services. Furthermore, the app provides real-time flight status updates, interactive airport maps, and travel assistance features, keeping travelers informed and well-equipped to navigate the airport with ease. Additionally, the centralized travel management platform allows users to effortlessly manage their travel itineraries, reducing the hassle of juggling multiple documents and reducing the risk of last-minute surprises. Overall, the NFC AIRPORT TOOL APP offers a unique solution to address the key pain points of air travel, making the airport experience more streamlined, secure, and personalized for travelers.

VI. REFERENCES

[1] S. Guizani, "Relay attacks concerns in wireless ad hoc, sensors, and RFID networks," Wire- less Commun. Mobile Comput., vol. 16, no. 11, pp. 1431–1435, Aug. 2016 .

[2] P. Pourghomi, M. Q. Saeed, and G. Ghinea, "A secure cloud-based NFC mobile payment protocol," Int. J. Adv. Comput. Sci. Appl., vol. 5, no. 10, pp. 24–31, 2014.

[3] A. Kumar, N. Saxena, G. Tsudik, and E. Uzun, "Caveat eptor: A comparative study of secure device pairing methods," in Pervasive Computing and Communications, 2009. PerCom 2009. IEEE International Conference on. IEEE, 2009, pp. 1–10

[4] Introduction to Secure Elements. Accessed: Feb. 9, 2021. [Online]. Available: https://globalplatform.org/wp-content/uploads/ 2018/05/Introduction-to-Secure-Element-15May2018.pdf

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

Signature Verification Using CNN and LPB

Ms.Anju T*[1], Gladia K.Lal[2], Hanna Shamsudheen[2], Jyothika Shaji[2], Nehamol Sunny[2]
*[1]Assistant Professor, Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Muvattupuzha, Kerala, India
[2]Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Muvattupuzha, Kerala, India

ABSTRACT
The efficacy of offline signature verification systems greatly depends on the feature extraction phase, as it directly impacts their ability to differentiate between genuine and forged signatures. This research introduces a novel hybrid approach that combines Convolutional Neural Networks (CNN) and Linear Binary Pattern (LBP) techniques for extracting features from signature images. In sectors such as banking, legal, and administration, where the authentication of handwritten signatures is crucial, offline verification systems play a vital role, emphasizing the importance of accurate feature extraction. Through the integration of CNN and LBP, followed by classification using SVM our method enhances the system's proficiency in precisely discerning between authentic and forged signatures. This underscores the pivotal role of feature extraction in optimizing the overall performance of the system.
Keywords: CNN, LBP, SVM

I. INTRODUCTION

Signature verification plays a pivotal role in numerous sectors, ranging from financial transactions to broader authentication systems. This project introduces an innovative hybrid methodology that exploits the synergies between Convolutional Neural Networks (CNN) and Local Binary Patterns (LBP) is employed for feature extraction. The fusion of CNN and LBP within this approach offers a potent and efficient solution for multi-classification tasks, with the overarching goal of augmenting the accuracy and reliability of signature authentication systems

The challenges faced by traditional signature verification methods lie in accurately distinguishing between authentic and forged signatures, attributed to variations in writing styles and the intricate nature of signature patterns. To overcome these challenges, the proposed hybrid method capitalizes on CNN's ability to extract intricate features from signature images and LBP's effectiveness in capturing shape and gradient information

Handwritten signatures function as distinct behavioural biometrics across various applications, including banking, financial documents etc. The verification of these signatures becomes particularly complex when dealing with unclear or ambiguous instances. Hence, the imperative for a system capable of effectively

discerning between genuine and fake signatures is vital to mitigate the risk of theft or fraud. Despite decades of research in this field, which has transitioned from traditional expert opinion-based verification to machine learning and contemporary deep learning algorithms, there remains a substantial need for the ongoing development and enhancement of offline signature verification systems.

A fundamental challenge in these systems lies in the precise extraction of discriminative features that can effectively differentiate between genuine and forged signatures. Conventional methods often depended on manually crafted features, which sometimes fell short of capturing the complex patterns and subtleties present in signatures. To address this limitation, the adoption of deep learning techniques has proven instrumental, given their capacity to autonomously extract hierarchical representations directly from raw data. This hybrid methodology aims to significantly enhance the discriminative power of the extracted features, ultimately contributing to the heightened accuracy and reliability of offline signature verification systems.

II. LITERATURE SURVEY

Several methods which help in signature verification have been developed recently. Many algorithms are employed to detect forged signature produced that are having a great effect on day today life.

A. OC-SVM and BC-SVM classifier

The focus is on recognizing the significance of handwriting as a unique biometric identifier, mainly in the context of signature verification for identity purposes. The researchers emphasize the demanding situations in developing universally proper algorithms for handwriting verification due to variations in patterns, even inside the identical man or woman. The observation compares the overall performance of OC-SVM and BC-SVM classifiers, aiming to beautify accuracy in figuring out solid signatures. The goals encompass determining the dominant capabilities influencing authenticity, comparing the effectiveness of the classifiers, and developing an OCR machine prototype for detecting forgeries. The look at scope is restricted to signatures using the English alphabet, with a dataset of 180 respondents elderly 15 to 60. The records, written on uniform white bond paper with particular ink and pen, are divided for schooling and trying out functions. The last aim is to make a contribution to the development of signature verification techniques with sensible programs in figuring out forgery [1].

B. Combinational Features and Secure KNN

This methodology encompasses four pivotal phases: feature retrieval, template synthesis, safeguarded preservation of templates and feature vectors, and the comparison process. The feature extraction procedure encompasses pre-processing, attribute creation, truncation, and quantization, alongside global and regional feature extraction. Templates are generated during the enrolment phase, and during comparison, the resemblance between the template and the feature vector derived from a test signature is assessed. To fortify security, both templates and feature vectors are shielded using a secure k-nearest neighbours (kNN) technique.

Prominent benefits of this approach include adept management of attribute truncation and quantization, comprehensive extraction of global and regional features for proficient signature representation, and the utilization of secure kNN for safeguarding templates and feature vectors while preserving verification

accuracy. However, there are potential hurdles involving the system's capability to tackle both spontaneous and skilled forgery. The former is approached as a matching challenge, while the latter is treated as a binary classification problem. Moreover, incorporating secure k-nearest neighbours (kNN) might result in increased computational demands, potentially impacting the verification process's efficiency. While the method presents a privacy-centric method for signature verification, featuring robust feature retrieval and secure data storage, it's essential to be mindful of the challenges associated with various types of forgery and the potential computational overhead in practical scenarios[2].

C. SVM and LBP

The study utilizes Local Binary Pattern (LBP) features, considered as pseudo-dynamic features capturing local properties of signature images. LBP features are chosen for their capability to differentiate genuine signatures from simulated forgeries, particularly when minor alterations are present. The novelty of the research lies in the exploration of LBP features within the WI framework for offline HSV, emphasizing their effectiveness in handling variations. Additionally, the research assesses the efficacy of utilizing the Support Vector Machine with Polynomial Kernel (SVM-POLY) classifier to construct an Offline Handwritten Signature Verification (HSV) system with writer-independent features. The objectives of the research include absorbing unfamiliar writers' signatures without retraining, minimizing the False Rejection Rate (FRR) through the SVM-POLY classifier, and assessing the effectiveness of LBP features and SVM-POLY classifier in the development of the offline HSV system[3].

D. Analysis of handwritten dynamic signature

It introduces a novel approach to verification of identity through the analysis of dynamically handwritten signatures, addressing the crucial issue in biometrics. The proposed method focuses on signature partitioning, uniquely utilizing hybrid partitions constructed through a blend of vertical and horizontal segments of the signature. These divisions correspond to distinct time intervals and variations in pen dynamics throughout the signing process. The algorithm strives to boost precision by selectively choosing and processing these amalgamated partitions, while also considering the stability of the signing process. The study employs two databases, MCYT100 and BioSecure, to assess the efficacy of the proposed method. The approach integrates fuzzy sets and fuzzy systems theory to gauge the comparison of test signatures with reference signatures. This yields individualized and precise results for each user. The paper contributes to the progress of dynamic signature analysis by furnishing a comprehensive depiction of the algorithm and showcasing simulation results obtained from the tested databases.[4].

III. PROPOSED SYSTEM

A. Architecture

The proposed system includes signature verification using convolutional neural networks (CNN) and local binary patterns (LBP), offering an intriguing technique that amalgamates deep mastering with conventional laptop vision techniques. To accomplish this, the number one priority is assembling a comprehensive dataset that encompasses both genuine and forged signatures. This dataset needs to be meticulously curated to consist of a variety of signatures, taking pictures of diverse writing patterns and versions. To ensure the version's robustness and generalization abilities, preprocessing steps are carried out on the collected photos.

This consists of resizing the signatures to a uniform format, normalizing pixel values to a standardized variety, and using statistical augmentation strategies. Techniques consisting of rotation, flipping, and scaling are used to enhance the dataset, improving the model's potential to understand signatures with various orientations and perspectives.

The aim of those pre-processing steps is to create a nicely organized and representative dataset that enables effective schooling and validation of the signature verification version. After completing the information series and pre-processing steps, the following stages contain statistics splitting, function extraction, and version improvement. The dataset is divided into schooling and trying-out sets; normally, 80% training and 20% testing are cut up to make certain an effective assessment of the model's overall performance. Local Binary Pattern (LBP) is used to seize texture features from signature photographs, exploiting nearby patterns to offer a valuable illustration for signature analysis. Simultaneously, a pre-trained convolutional neural network (CNN), together with VGG16 or ResNet, is employed for characteristic extraction.

The CNN is first-rate-tuned at the signature dataset, adapting its parameters to the unique assignment of signature verification. The functions extracted from LBP and CNN are fused through concatenation, developing a comprehensive feature representation for every signature. This fused illustration serves because it enters a neural community, usually comprising fully related layers, designed to examine the problematic courting between these features and the authenticity of the signature. The model is skilled in the usage of the school dataset, utilizing appropriate loss features like binary cross entropy and optimization algorithms. To validate the version's generalization competencies and protect against overfitting, a validation set is hired, bearing in mind the excellent tuning of hyperparameters. Once the education and validation phases are complete, the version undergoes evaluation of the usage of the testing dataset, with metrics like accuracy, precision, F1 rating, and recall presenting a complete evaluation of its effectiveness.

Classifier— SVM (Support Vector Machine) tests the accuracy. Hyperparameter tuning follows, enhancing the model's performance with the aid of refining its configuration. Upon achieving first-class performance, the version is deployed for real-international signature verification responsibilities. To ensure ongoing efficacy, a strategy for non-stop development is applied, related to ordinary updates and retraining with new information. This iterative manner aims to preserve the model as adaptive and strong over the years, addressing the evolving challenges of signature verification in sensible packages.

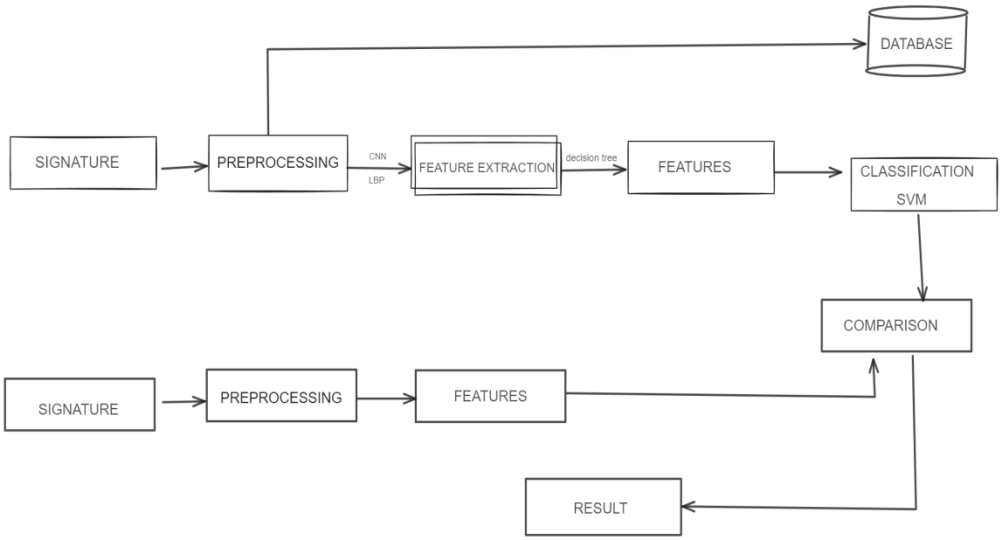


Fig-1: Architecture diagram

B. Implementation

The implementation of the proposed Signature verification model involves an innovative hybrid feature extraction approach by combining CNN and LBP. This methodology aims to enhance the system's ability to distinguish between genuine and forged signatures. The process involves separate feature extraction using CNN and LBP. The VGG16 model is used for feature extraction .The CNN training depends on a threshold value which is set to 0.9 based on the false positive and false negative. The threshold value can be changed according to the value of false positive and false negative. The features extracted from both CNN and LBP are then merged. Evaluation is conducted using one classifier - SVM to assess the overall effectiveness of the hybrid method. The dataset consists of real and forged signatures of equal numbers. A total of 16,000 signatures are used for training purposes. The model undergoes training with fine-tuned CNN parameters and validation to ensure generalization. Performance measures such as F1 score, precision, recall and accuracy are utilized for rigorous evaluation. The integration of CNN and LBP facilitates a comprehensive feature extraction strategy, capturing both global and local characteristics from signature images, contributing to heightened accuracy in signature verification tasks

REFERENCE	TITLE	TECHNIQUE	MERITS	DEMERITS
[1]	Offline Handwritten Signature Verification Using OC-SVM And BC-SVM Classifier	• OC-SVM • BC-SVM	• Involve testing for quality • High accuracy rate	• Requires proper identification of features • Detect variations in signature written by the same person as forged
[2]	A Privacy-Preserving Handwritten Signature Verification Method Using Combinational Features and Secure KNN	• Combinational features • Secure KNN • SVM	• Data Confidentiality • Protect user template and features	• Computational overhead • Complexity
[3]	Offline Signature Verification Using Support Local Binary Pattern	• SVM • LBP	• Classify signature without retaining • Simulation investigation	• Complexity of polynomial kernel • Limited database size

[4]	A New Algorithm for Identity Verification based on the Analysis of a Handwritten Dynamic Signature	• SVM • ANN • HMM	• Use fuzzy sets and fuzzy systems theory • Determines values of weights	• Sensitive to changes of a handwriteen signatures • Mechanism of the similarity evaluation is complex

Table 1 : Comparative Study

IV. CONCLUSION

In conclusion, the integration of convolutional neural networks (CNN) and local binary patterns (LBP) for signature verification represents a promising approach to enhance the precision and dependability of the authentication procedure. The proposed method leverages the power of CNN to learn hierarchical features from signature images, capturing intricate patterns and variations. Additionally, the incorporation of LBP provides valuable texture information, contributing to the system's ability to discern between authentic and forged signatures. The feature selection algorithm, rooted in decision trees, further fine-tunes the extracted features, emphasizing crucial discriminative elements. The combination of CNN and LBP is complemented by the inclusion of a comprehensive feature extraction strategy. The classification approach employs one classifier— Support Vector Machine (SVM) to assess the efficacy of the hybrid method.

V. REFERENCES

[1] Mary Jane C. Samonte, Roxanne Michelle G.Eullo, And Alan I.Misa School of Information Technology Mapua University Makati City,Offline Handwritten Signature Verification Using OC-SVM And BC-SVM Classifier,Philippines,2017

[2] Z. Xia, T. Shi, N. N. Xiong, X. Sun and B. Jeon, "A Privacy- Preserving Handwritten Signature Verification Method Using Combinational Features and Secure KNN," in IEEE Access, vol.6, pp. 46695-46705, 2018, doi: 10.1109/ACCESS.2018.2866411.

[3] P, Vickram& A, Sri &Swapna, Dheeravath. (2016). Offline Signature Verification Using Support Local Binary Pattern. International Journal of Artificial Intelligence & Applications. 7. 85-94. 10.5121/ijaia.2016.7607.

[4] K. Cpałka, M. Zalasinski, and L. Rutkowski, "A new algorithm for ′ identity verification based on the analysis of a handwritten dynamic signature," Appl. Soft Comput., vol. 43, no. 1, pp. 47–56, June. 2016. c

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

Beyond Words : Classifier-B's Journey to Deeper Understanding

Ms. Manjusha Mathew*[1], Christo Robinson[2], Joel Jose[2], Pranav T Pramod[2], Sidharth S[2]
*[1]Assistant Professor, Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Muvattupuzha, Kerala, India
[2]Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Vazhakulam, Kerala, India

ABSTRACT
In the ever-expanding universe of data, organizations grapple with the challenge of effectively classifying and extracting insights from diverse formats, including documents, images, and videos. Traditional methods often fall short, hindered by manual effort or limited capabilities. Enter" Classifier -B," a software solution that aims to revolutionize data classification by harnessing the power of NLP. Classifier-B is a marvel of modern technology, blending the best of different fields to create a unified understanding of diverse data. At its core, it employs BERT, a state-of-the-art NLP model, to delve into the semantic depths of text, deciphering meaning and intent with exceptional accuracy. This allows for precise classification of documents based on their content, sentiment, and linguistic features. However, the software's capabilities extend beyond the written word. By extracting text from images and video transcripts, Classifier -B expands its classification prowess to encompass the rich insights embedded within visual media.

Keywords - Natural Language Processing, Bidirectional Encoder Representations from Transformers, Optical Character Recognition

I. INTRODUCTION

Information nowadays bombards us from each course, a great ocean of files, pictures, and movies pulsating with what it means. Yet, extracting that, which means reworking facts into actionable insights, remains a monumental undertaking. Traditional equipment often conflicts, clinging to singular facts, like islands in a hurricane, failing to comprehend the full picture. This venture sets sail on a distinctive course, charting a path towards a brand-new horizon i.e. A software program answer that harnesses the strength of a NLP model to navigate this sea of facts with unequalled clarity and precision.

Imagine an economic report, its crisp paragraphs punctuated by way of charts and graphs that whisper a deeper narrative. Relying completely on text analysis might be akin to analysing the simplest headings on a map, oblivious to the vast terrain the symbols depict. Our software transcends such obstacles, embracing the symphony of record types. It makes use of the nuanced expertise of textual content by using the latest NLP model, BERT, to deconstrue the written word with surgical precision. Simultaneously, advanced laptop

imaginative and prescient algorithms delve into the visible realm, decoding the stories woven into pixels and the feelings etched in frames. And while confronted with textual content embedded inside the visual tapestry, the magic of optical person recognition bridges the gap, transforming those silent symbols into eloquent voices.
By orchestrating this harmonious interaction of technology, our software achieves a profound feat. It comprehends the sector now not through fragmented glimpses but through a holistic lens. It analyses files, images, and videos in tandem, unearthing connections invisible to normal model approaches. This unlocks a brand-new generation of class accuracy, empowering us to prepare, categorize, and apprehend information with unparalleled performance and depth. This journey transcends mere technical advancement; it holds the capacity to revolutionize the way we interact with the arena around us, from streamlining record control to unlocking deeper insights from multimedia content across diverse domains. While BERT masters the difficult nuances of language, our software program recognizes that the tale does not end with the written phrase. Images and movies, too, speak their very own dialect, brimming with visible cues and dynamic narratives. To decipher this particular language, we equip our software with state-of-the-art photograph and video-type techniques.This is simply the beginning of our voyage. The future beckons, promising the growth of our statistical kinds, the non-stop refinement of our algorithms, and the advent of consumer-friendly interfaces that democratize this notable strength. Ultimately, our ambition is to build an international system wherein machines can definitely recognize the arena in all its multi-version richness, guiding us in the direction of a future where records dance, not drown, in the considerable ocean around us. The implications of this software programme go far beyond the classroom. By unlocking the latent know-how inside data, we empower businesses to gain a deeper understanding of their customers, optimize internal approaches, and make information-driven decisions with greater self-belief. The software acts as a Rosetta Stone, interpreting the hidden language of data and remodelling it into a powerful tool for strategic advantage. As we embark on this adventure of innovation, we stand at the brink of a brand-new era in which statistics is not an uncharted desert but a navigable galaxy ready to be explored and its treasures unearthed.

II. RELATTED WORKS

A. BERT and GA

Identifying the appropriate category for documents is essential for green fact retrieval, organisation, and analysis. Yet, conventional methods like cosine similarity and BM25 often stumble. Their reliance on simplistic bag-of-words processes overlooks the tricky semantic connections and dependencies within documents, leading to misclassifications for nuanced subjects or specialised vocabularies. To triumph over those boundaries, Mustafa et al. (2023) endorses a progressive approach leveraging genetic algorithms (GAs) [1]. Inspired by natural choice, GAs evolve a populace of candidate solutions (document characteristic units) through generations. Each" chromosome" represents a particular function aggregate, and its" fitness" is decided by how as it should be, it classifies schooling records files. Over successive iterations, the fittest chromosomes (feature sets) are replicated and similarly optimized through choice, crossover, and mutation. This evolutionary method gradually sculpts tremendously correct classification fashions. To quickly take a look at their method, Mustafa et al. turned to the wealthy ACM dataset compiled by Rodrigues and Santos (2009). These numerous collections of computer technological know-how research articles, richly annotated with study areas and keywords, supplied a robust platform to assess the GA's effectiveness in a complicated area and the effects had been

groundbreaking. GAs excels at navigating the difficult semantic and stylistic relationships inside files, resulting in extra-nuanced and correct classifications. GAs without problems adapt to address massive volumes of facts correctly, making them a promising answer for actual global facts processing tasks. Utilizing GAs, Mustafa et al. achieved tremendous improvements in class accuracy compared to traditional methods.

B. BERT and CNN

Chen et al. (2022) [2] advocate a groundbreaking approach that harnesses the mixed energy of powerful tools: BERT and CNN. BERT, a pre-trained language version, shines a mild light on the context of words, interpreting their complex relationships inside sentences. CNN, a master of pattern popularity, excels at spotting neighbourhood functions and hidden patterns inside textual content. This dynamic duo works in ideal harmony. BERT generates contextualized phrase embeddings, allowing each phrase to be understood when it comes to its surrounding context. These embeddings are then fed into a multi-filter-out CNN structure. Smaller filters act like magnifying glasses, zooming in on nearby features like n-grams, whilst larger filters take a much wider view, taking pictures with longer-range dependencies and sentence-level styles. The blended output, a rich tapestry of nearby and global capabilities, empowers the version to make accurate classifications with newfound self-belief. This dynamic duo works in ideal harmony. BERT generates contextualized phrase embeddings, allowing every phrase to be understood with regards to its surrounding context. These embeddings are then fed right into a multi-clear out CNN structure. Smaller filters act like magnifying glasses, zooming in on neighbourhood capabilities like n-grams, while large filters take a much broader view, capturing longer-range dependencies and sentence-stage patterns. The blended output, a wealthy tapestry of local and worldwide capabilities, empowers the model to make correct classifications with newfound confidence. To put their introduction to the test, Chen et al. enlisted formidable datasets: THUCNews, a substantial collection of Chinese information articles classified into 10 awesome topics, and CLUE, a numerous library brimming with articles throughout various classes. When pitted in opposition to traditional strategies and different deep mastering models, the BERT-CNN technique emerged positive, achieving considerably better accuracy on both datasets.

C. MLM and CLOZER

Open cloze questions (OCQs), in which a word is intentionally unnoticed from a sentence and the learner must fill it in, are like linguistic puzzles unlocking comprehension, important thinking, and vocabulary improvement. They're especially robust for assessing and nurturing language studying. But crafting exceptional OCQs demands sizable linguistic information and time, proscribing their use and hindering customized mastering. Matsumori et al. (2023) [3] gift CLOZER, a modern ML-powered solution that automates OCQ technology, releasing educators and language learners from the shackles of guide efforts. CLOZER leverages the magic of Masked Language Models (MLMs) like BERT, educated on good sized amounts of text statistics to expect likely phrase sequences. CLOZER's prowess changed into placed to the take a look at on 1, six hundred sentences from various educational texts, each paired with a pre-written OCQ and solution CLOZER stands as a testimony to the electricity of AI in revolutionizing schooling. With similarly research and development, its algorithm can be subtle, and its integration into getting to know platforms explored, paving the way for a destiny in which OCQs liberate the total potential of language novices and educators alike.

D. Keyword Extraction

Extracting key concepts from text is the compass to navigate this information overload. But traditional methods like TextRank and TF-IDF, blinded by mere word counts, often miss the deeper connections and context. Irrelevant keywords and overused terms abound, leaving us adrift in a frustrating ocean of text. Xu et al. (2021) [4] present a groundbreaking approach that marries word frequency with word association, revealing a hidden map of both prominent and semantically relevant keywords. We start with the established TF-IDF score, a beacon illuminating a word's importance within a document and the broader corpus. This ensures frequent and relevant terms rise to the top, ensuring we don't miss the obvious landmarks The authors extract keywords using both frequency and association independently, then unveil the terms that appear on both lists. This intersection ensures keywords are not only statistically prominent but also semantically connected, forming a cohesive and insightful treasure map. To prove its adaptability, this method is tested on two diverse landscapes: the vast ACM Digital Library corpus and a collection of focused scientific abstracts. This demonstrates its effectiveness across different domains and text lengths, from comprehensive databases to concise research summaries. This approach is more than just a treasure map; it's a new way to navigate the textual jungle. Imagine research databases that surface hidden gems based on deeper connections, or news summaries that capture not just trending topics but also underlying themes and relationships. This is the promise of a future where the key to overload lies not just in quantity, but in quality and connections. ECHR Classification.

E. ECHR Classification

Navigating the complex landscape of legal documents can be a daunting task, especially when it comes to pinpointing the specific human rights violations at play. Imran et al. (2023) [5] take aim at this challenge, unveiling a groundbreaking approach that leverages the power of transformer-based language models to untangle the legal jargon and identify the crux of the matter. Classifying cases from the European Court of Human Rights (ECHR) based on the type of violation involved is no easy feat. Lengthy, intricate documents, teeming with legalese, often obfuscate the core issues. Traditional methods struggle to decipher the nuances and grasp the context, leaving researchers and legal professionals lost in a maze of paperwork. To illuminate the path, Imran et al. turn to a team of formidable champions: BERT, RoBERTa, ELECTRA, and DeBERTa. These transformer-based models, masters of extracting meaning from complex text, possess the keen eyes needed to dissect legal prose. Pre-trained on vast datasets like BooksCorpus and English Wikipedia, they are
further honed on a smaller, targeted collection of ECHR judgments meticulously labelled with specific violations.

Imran et al. (2023) offer a tantalizing glimpse into a future where technology empowers legal scholarship and enhances access to justice. With the power of transformers guiding the way, unravelling the legal labyrinth and ensuring human rights are upheld becomes a more attainable mission. Let's embrace this technological transformation and pave the path for a more transparent and efficient legal system, one classified case at a time.

F. MEDLINE and CREGX

Delving into the sector of biomedical studies, correct textual content type is the important thing that unlocks expertise and empowers clinical selection-making. However, constructing efficient classification fashions often stumbles upon a familiar obstacle: the scarcity and fee of manually categorized information. Labelling extensive collections of medical texts demands meticulous expertise and precious time, slowing down progress and

hindering version performance. Flores et al. (2021) [6] unveil a captivating solution – a lively gaining knowledge of (AL) technique that taps into the strength of routinely generated everyday expressions (CREGEX). Imagine this CREGEX as a tireless detective, scouring pre-classified documents for treasured clues – ordinary styles and linguistic fingerprints that signal particular biomedical principles. With every file analysed, CREGEX refines its understanding of these clues, turning into increasingly more adept at identifying them in unlabelled texts. AL comes to the rescue with a strategic query approach. Analysing the unlabelled files, CREGEX identifies those which are most" uncertain" – texts in which it hesitates in assigning a class. These uncertain documents hold the finest capability for studying, as labelling them famous new clues and allows CREGEX to refine its class abilities. Flores et al. (2021) display the efficacy of their CREGEX-AL method on two essential biomedical text datasets – MEDLINE abstracts on diabetes and PubMed Central articles on breast cancer. Beyond the magnificent accuracy gains, CREGEX gives another precious benefit: transparency. Unlike black-box fashions, CREGEX's common sense is laid bare thru its without problems interpretable ordinary expressions. This transparency fosters trust and lets in researchers to delve deeper into the version's reasoning, building self-belief in its conclusions.

III. COMPARITIVE STUDY

The assessment of existing technologies plays a huge role in developing new project. In Fig. 1. We analyse the merits and demerits of the existing technologies mentioned in the related works

Reference	Title	Merits	Demerits
1	Optimizing Document Classification: Unleashing the Power of Genetic Algorithms	•Can explore large datasets. •Can be customized to fit specific document classification datasets.	•Expensive. •Sensitive to choice of parameters. •Slow to arrive at solution.
2	Exploring Natural Language Processing in Model-To-Model Transformations	•NLP resolves ambiguity and inconsistencies. Reduce manual efforts and improve efficiency	•Expensive. •Insufficient data can cause errors •Doesn't support cross datasets
3	Extracting Keywords from Texts based on Word Frequency and Association Features	•Simple and easy. •Effective in scientific papers and technical documents.	•May not be effective for other datasets •Sensitive to parameters •Data sufficiency.
4	Deep learning-based text knowledge classification for whole-process engineering consulting standards	•Highly accurate •Can use large datasets	•Complex to train and deploy •Require lots of data •Ambiguity on decisions of model.
5	Mask and Cloze: Automatic Open Cloze Question Generation Using a Masked Language Model	•Easy •Educational purposes •High quality	•Large data needed •May not be suitable for every datasets •Needs further human evaluation
6	A Long-Text Classification Method of Chinese News Based on BERT and CNN	•High accuracy on dataset •Can handle long data •Easy to train	•Require large data •Expensive •Difficult to debug •Cant use other datasets

Fig. 1. Comparison of technologies

IV. PROPOSED SYSTEM

A. Architecture

In the proposed system, Classifier-B, we are implementing a BERT based text classification. Utilizing BERT for classifying various data such a text, image, video, our system tries to identify and classify various types of data and their content. It first gathers your data, be it text-heavy documents, images, or insightful videos. Preprocessing ensures everything is ready for analysis, including extracting text from visual media using OCR technology and tag extraction. Under the hood, BERT model, analyse the content, deciphering meaning and intent, before assigning the most appropriate category based on topic, sentiment, or defined criteria. The results then take centre stage on an interactive interface, ready for exploration and further action. With Classifier-B, understanding your data becomes a harmonious experience, unlocking valuable insights from every corner of your document universe. The full software is built as a webapp using PHP. The architecture of the proposed system is given in Fig.2.

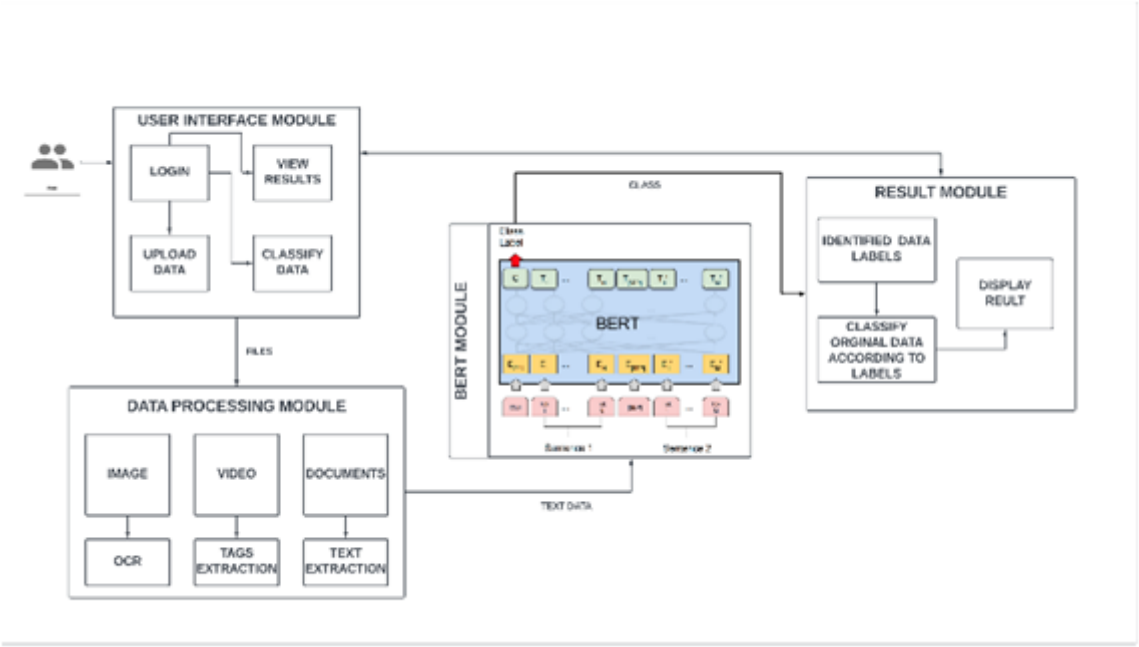


Fig.2.Architecture of Classifier-B

The diagram depicts the journey of data, starting from the user. It undergoes appropriate processing measure, which is different for each file types before being refined further through stemming/lemmatization and stop word removal. This prepped data is then fed into the BERT model, which in turn makes predictions based on the learned patterns from training. In short, raw data is transformed, fed to a learning model, and emerges as meaningful predictions.

B. Implementation

User Interface Module: The module will receive data in the form of docs, image, video from the users. To begin the process of classification, it is necessary to gather the well-made data that you are interested in classifying. This can be done by selecting data from a specific and valid source. Once the desired data have been identified and collected, they can then be analysed and pre-processed using various methods or techniques in order to draw insights or conclusions about the type of content.

The UI allows the user to login, upload specific data files, classify the data, and view the results of classification.
Data Processing Module: This module prepares the data for analysis by the BERT model. Processing the data involves using various techniques for each file type. To extract the text from images we utilize the OCR (optical character recognition), To extract subtitles from video we use YouTube data API. Documents files are converted to txt files and then fed into the model. Data preprocessing methods to check data validity and type checking are implemented along with methods to detect corrupted files. Other ways involve tasks such as removing unnecessary characters or words, correcting spelling errors, stemming or lemmatizing words to their base form, or creating numerical representations of the documents and text through BERT's own encoder. The goal of preprocessing is to clean and organize the data in a way that makes it more useful and effective for the next steps in the process.
BERT Module: In the classification and prediction phase of using BERT for text classification, the pre-processed input text is passed through the fine-tuned BERT model. As the text traverses through the layers of the model, it undergoes transformations that capture complex linguistic patterns and contextual information. These transformations allow the model to generate rich representations of the input text. At the end of the model, these representations are fed into additional task-specific layers, which have been trained to map the learned representations to the target classification labels. The output of these layers is a probability distribution over the possible classes, obtained through SoftMax activation, indicating the likelihood of the input text belonging to each class. The class with the highest probability is selected as the predicted class for the input text. This process allows BERT to effectively classify text based on the patterns and features it has learned during fine-tuning, making it a powerful tool for various natural language processing tasks.
Result Module: This module presents the outcome of the classification process to the user. It involves displaying the classified data with assigned categories, along with any other previous classification details and data. The result section is accessible through the UI and users can easily navigate through the interface.

V. CONCLUSION

"Classifier -B" represents more than just a software solution; it signifies a pivotal moment in the evolution of data intelligence. It heralds a new era of mastery over information for businesses. No longer constrained by isolated data silos or laborious manual methods, the path to actionable insights now lies through the harmonious orchestration of knowledge enabled by "Classifier -B." Its benefits transcend mere classification accuracy. Acting as a catalyst for knowledge discovery, this software empowers organizations to unearth previously concealed treasures within their existing data repositories. Picture ancient images recounting their stories, product packaging offering personalized guidance, and video footage unveiling unexpected patterns of consumer behaviour—all illuminated by the prowess of Natural Language Processing (NLP). "Classifier -B" stands as a testament to the relentless pursuit of innovation in the realm of data. It symbolizes a future where businesses flourish not solely by the quantity of data they possess, but by their capacity to unlock its true potential. As we venture further into this transformative era, "Classifier -B" serves as a guiding beacon, illuminating the path toward a data-driven future ripe with unexplored opportunities.

VI. REFERENCES

[1] Mustafa, Ghulam, et al. ”Optimizing document classification: Unleashing the power of genetic algorithms.” IEEE Access (2023).

[2] Chen, Xinying, Peimin Cong, and Shuo Lv. ”A long-text classification method of Chinese news based on BERT and CNN.”IEEE Access 10(2022):34046-34057.

[3] Matsumori, Shoya, et al. ”Mask and Cloze: Automatic Open Cloze Question Generation Using a Masked Language Model.” IEEE Access 11 (2023): 9835-9850

[4] Xu, Zhenzhen, and Junsheng Zhang. ”Extracting keywords from texts based on word frequency and association features.” Procedia Computer Science 187 (2021): 77-82.

[5] Imran, Ali Shariq, et al. ”Classifying European Court of Human Rights Cases Using Transformer-Based Techniques.” IEEE Access (2023).

[6] Flores, Christopher A., Rosa L. Figueroa, and Jorge E. Pezoa ”Active learning for biomedi-cal text classification based on automaticallygenerated regular expressions.” IEEE Access 9 (2021): 38767-38777.

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

An Examination on Credit Card Fraud Detection using Machine Learning

Ms.Swathi Venugopal*[1], Clare Maria Francis[2], Gopika Byju[2], Mariya P B[2], Anns K James[2]
*[1]Assistant Professor, Department of Computer Science, VJCET, Ernakulam, Kerala, India
[2]Department of Computer Science and Engineering, VJCET Ernakulam, Kerala, India

ABSTRACT

The widespread use of credit cards for online transactions has led to increased efficiency and user convenience. However, this surge in usage has also amplified the risk of misusing credit card . Frauds associated with credit cards result in substantial losses in finance for both cardholders and financial institutions. This research focuses on detecting such frauds, considering challenges like accessing public data, dealing with high-class imbalance data, evolving fraud patterns, and addressing high false alarm rates. Existing literature explores various machine learning methods for credit card fraud detection, such as Extreme Learning Method, Decision Tree, Random Forest, Support Vector Machine, Logistic Regression, and XG Boost. Despite these efforts specifically low levels are available, so state- of-the-art deep learning algorithms are needed to reduce fraud losses.

This study conducts a comparative study of machine learning and deep learning algorithms, using the European card bench- mark dataset for fraud detection. Initially, a machine learning algorithm is applied, yielding improved fraud detection accuracy. Subsequently, three convolutional neural network architectures are employed, progressively enhancing detection performance through the addition of layers. The empirical analysis involves differences in the number of layers stored, epoches, and use of the latest models. Analysis of the research project reveals remarkable improvements in accuracy, f1-score, accuracy, and AUC curves , with optimal values of 99.9, 85.71, . 93, and 98, respectively. The proposed model outperforms machine learning and deep learning algorithms for credit card redemption. Further experiments, including data balancing and the use of deep learning algorithms, will focus on reducing the false negative rate, making the proposed methods more effective in the world real internal credit card fraud detection.

I. INTRODUCTION

The digital revolution has made credit cards ubiquitous, fa- cilitating transactions and providing unparalleled convenience. However, this ease of use has also provided a distracting target for malicious people, leading to a worrying increase in unauthorized card usage. These fraudulent transactions have severe financial losses on those who are robbed of their hard- earned money, and on the financial institutions that bear the burden of fees and reputational damage.

For years, financial institutions have relied on traditional methods such as rules-based systems and audit analysis to combat this threat. But as fraudsters take more sophisticated tactics, these tactics struggle to keep up. This is where the transformative power of machine learning (ML) enters the scene. Using big data of historical transactions as its weapon, ML empowers algorithms to identify complex patterns and anomalies that often indicate fraudulent activity Unlike traditional methods, ML is incredibly powerful constantly adapting to new data and changing fraud trends, enabling robust real- time detection.

This study explores the fascinating world of ML techniques for credit card fraud detection, focusing on a power- ful combination: SMOTE (Synthetic Minority Over-sampling Technique) and AdaBoost (Adaptive Boosting) to identify algorithms that most effectively possible through rigorous research, data -We aim to demonstrate the utility of SMOTE in solving the imbalance challenge, and ultimately promote the use of potentially optimal systems will propose a solution to this widespread problem.

The implications of this research are multifaceted. It safe- guards hard-earned personal funds by empowering financial institutions to detect and prevent fraudulent transactions and strengthens the overall security of the digital financial environment. Additionally, this study provides valuable insights to researchers seeking to improve fraud detection, ultimately building trust and confidence in the digital financial system.

This revised edition builds on the strengths of the original text to provide additional information and insight: Emphasizes the paradoxical aspects of the security and risks associated with credit cards. Extends the limitations of traditional methods approach to fraud detection. Highlights the unique capabilities and flexibility of ML .Provides clarity description of selected ML strategies and potential benefits. Financial institutions only but shows wider implications of the study.

II. LITERATURE SURVEY

A. ADABOOST AND SMOTE

Credit card fraud detection poses the most important chal- lenge due to the inherent imbalance in transaction data, where fraudulent instances are vastly outnumbered by legitimate ones. This imbalance can significantly hinder the performance of traditional machine learning algorithms, leading to missed detections and financial losses. This paper addresses this problem by investigating the effectiveness of two techniques: SMOTE (Synthetic Minority Over-sampling Technique) and AdaBoost, to increase the efficiency of machine learning tech- niques for detecting credit card fraud as shown in the figure

- The research paper[1] aims to evaluate the effectiveness of these techniques, individually and in combination, on various popular machine learning algorithms and identify the optimal approach for accurate and robust fraud detection amidst im- balanced data. It dives into the realm of credit card fraud detection, tackling the inherent challenge of imbalanced data where fraudulent transactions are scarce. To address this, the authors embark on a comprehensive evaluation, employing two key techniques: SMOTE (Synthetic Minority Over-sampling Technique) and AdaBoost.

The foundation lies in the ”European cardholders dataset,” representing real-world transactions. However, the imbalanced nature, with far fewer fraudulent cases, poses a hurdle for traditional machine learning algorithms. To overcome this, SMOTE steps in, synthetically generating new ”fraudulent” instances based on existing ones.This aims to create a more balanced dataset, ensuring both legitimate and fraudulent transactions have equal footing during model training.The study of paper[1] deploys a diverse set of six algo- rithms: Support

Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), Extreme Gradient Boosting (XG- Boost), Decision Tree (DT), and Extra Tree (ET) as shown in figure 1. Each brings its own strengths and weaknesses to the table, offering a variety of perspectives on fraud detection.

To assess the performance of these algorithms, a battery of metrics is employed. Accuracy, recall, precision, Matthews Correlation Coefficient (MCC), and Area Under the Curve (AUC) provide a comprehensive picture. Accuracy takes all accuracy, while recall and accuracy go deeper and deeper into detecting deceptive tasks (recall) and avoiding false alarms (accuracy).MCC offers a balanced view, and AUC summarizes the ability of model to rank fraudulent transactions correctly.AdaBoost, an ensemble learning technique, enters the arena. It trains multiple ”weak learners” (base algorithms) iteratively, focusing on instances misclassified by previous learners. These learners are then combined into a stronger ”en- semble” model, aiming to surpass the individual performance of each base algorithm.

It meticulously evaluates the impact of SMOTE and Ad- aBoost, both individually and in combination (SMOTE- AdaBoost). This allows for understanding how each technique influences the performance of different algorithms and how their combined effect compares.The research in paper [1] aims not only to demonstrate the best methodology for credit card fraud detection on imbalanced datasets but also to provide valuable insights into the individual and combined impact of SMOTE and AdaBoost achieve on machine learning algorithms. A crucial challenge in the financial domain: imbalanced data. To train effective fraud detection models, the authors leverage a specific dataset. The dataset hails from European cardholders, aiming to provide real-world, practical context for the study. This ensures the relevance of the findings to actual credit card transactions and their inherent patterns.

The core issue lies in the class imbalance: legitimate trans- actions vastly outnumber fraudulent ones. This often presents models that has biased patterns for most people, which can miss subtle red flags of fraud. The paper indicates a total of 284,807 transactions, with only 0.172 percentage being fraudulent. This translates to a highly imbalanced distribution, emphasizing the need for techniques like SMOTE to address this problem effectively. By analyzing the characteristics of this dataset, we gain valuable insights into the real-world scenario of credit card fraud detection. It evaluates various approaches to combat the critical challenge of imbalanced data in fraud detection.

Applying SMOTE significantly improved the overall performance of all machine learning algorithms, particularly in terms of recall (identifying fraudulent transactions).Among the evaluated algorithms, Extreme Gradient Boost- ing (XGBoost) emerged as the frontrunner, achieving the highest accuracy, recall, precision, and AUC, both with and without SMOTE-AdaBoost. This highlights how effective the XGBoost is at detecting credit card fraud.

The combined application of SMOTE and AdaBoost (SMOTE-AdaBoost) consistently outperformed individual techniques, demonstrating its effectiveness across various eval- uation metrics and algorithms. This approach achieved the highest overall recall (identifying 98.79 percentage of fraud- ulent transactions), crucial for minimizing financial losses. Testing on a synthetic, highly skewed dataset confirmed the generalizability of the findings, solidifying the robustness of SMOTE-AdaBoost across different data distributions. This adds confidence to the identified optimal approach. These results offer valuable insights for credit card fraud detection systems. By leveraging SMOTE to address imbalanced data and employing AdaBoost to enhance ensemble learning, orga- nizations can achieve more accurate and robust fraud detection, safeguarding financial security and minimizing losses.

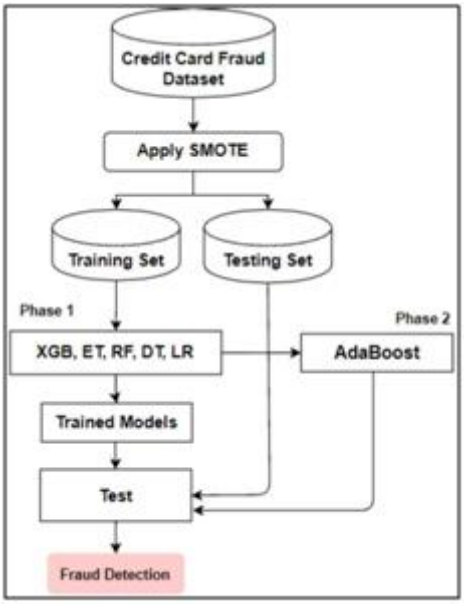


FIGURE 1:Framework

In figure 1 the framework of the system using mainly adaboost and smote algorithms along with RF,LR is shown.The frame- work consists of four main stages: data preprocessing, feature selection, model training, and model evaluation.

B. ADABOOST AND MAJORITY VOTING

The exponential rise of credit card fraud demands sophisti- cated detection systems, as traditional methods falter against evolving strategies and imbalanced data.This research paper[2] explores the potential of combining AdaBoost, a powerful ensemble learning technique, and majority voting, a robust ensemble selection method, to tackle this challenge. We aim to determine if this novel approach can outperform conventional single classifiers in accurately pinpointing fraudulent transac- tions, while addressing the inherent class imbalance in credit card data. We want to study how different base classifiers and methods of majority voting affect performance metrics such as true positive rate and false positive rate. Our goal is to understand how well this combined approach works in creating a strong and precise system for detecting credit card fraud. Ultimately, our aim is to help financial institutions avoid significant losses and maintain trust with their customers.

To tackle credit card fraud, this research paper[2] em- ploys a two-pronged methodology leveraging the strengths of AdaBoost and majority voting. At the start, we assess twelve varied machine learning algorithms, including tradi- tional neural networks and deep learning models, using both standard benchmark datasets and real-world datasets. This comprehensive assessment establishes a baseline performance landscape and identifies suitable base classifiers for AdaBoost integration.Next, AdaBoost is employed as the primary ensem-ble learning technique. Its iterative boosting process focuses on training successive weak learners on datasets weighted accord- ing to their misclassification errors in previous rounds. This adaptive approach gradually strengthens the ensemble's ability to identify both typical and outlier transactions. To further refine the prediction, majority voting is implemented as the ensemble selection method. Every base classifier contributes a vote for the class of a transaction (fraudulent or legitimate), and the final prediction relies on the majority decision. This straightforward but powerful technique diminishes the impact of weaknesses in individual base classifiers,

enhancing the overall reliability of the system. To assess how well this method works, we use a two-stage validation process. First, the performance of AdaBoost with majority voting is compared against individual base classifiers and other ensemble methods on the chosen datasets.We use metrics such as accuracy, precision, recall, and F1-score to measure how well fraudulent transactions are detected while keeping false positives to a minimum. In the second stage, we evaluate the model's resilience by introducing artificial noise into the real-world dataset. This simulates real-world scenarios where fraudulent activities may involve obfuscation techniques. By analyzing the model's performance under these conditions, the research aims to ensure its generalizability and effectiveness in practical applications. With this thorough approach, the study aims to validate the effectiveness of AdaBoost and majority voting in addressing credit card fraud, leading to the advancement of more precise and resilient fraud detection systems tailored for financial institutions.

The dataset utilized in this research paper[2] concerning credit card fraud detection is diverse, with the intention of offering a thorough assessment of the proposed AdaBoost and majority voting method. It employs a two-pronged strategy which are Public Benchmark Datasets that are readily available datasets that provide a standardized testing ground for com- paring the performance of our methodology against existing algorithms. Examples could consist of datasets provided by the UCI Machine Learning Repository or Kaggle competitions, specifically tailored for tasks related to detecting credit card fraud. Another source is real-world data from financial in- stitutions, which provides access to anonymized information to protect privacy. This data offers valuable insights into the unique challenges and intricacies of genuine credit card fraud patterns. This data often comprises millions of transactions with rich feature sets incorporating cardholder information, transaction details, and behavioral parameters. The research leverages both types of datasets for a holistic evaluation. Benchmark datasets offer generalizability and standardized comparisons, while real-world data provides domain-specific insights and ensures the practical applicability of the findings. Additionally, injecting artificial noise into the real-world data simulates the evolving tactics of fraudsters, further strengthen- ing the robustness of the model. By analyzing the performance of AdaBoost and majority voting on both public and real-world datasets, the research aims to establish their viability for real- world credit card fraud detection systems.

The experiment employing AdaBoost and majority voting for credit card fraud detection produced encouraging out- comes, demonstrating its potential to surpass conventional methods. Notable findings include enhancements in detection accuracy, improved handling of class imbalances, insights into the impact of base classifiers, and robustness to noise. Overall, the results strongly endorse the efficacy of employing AdaBoost and majority voting for detecting credit card fraud. This approach presents a promising path for constructing more precise and robust fraud detection systems, ultimately ensuring the protection of financial institutions and maintaining customer trust.

C. RANDOM FOREST AND SVM

In the domain of monetary protection, identifying fraud activities in credit card transactions is a top priority, particularly with the rapid expansion of online commerce. This underscores the need to create resilient and effective fraud detection systems capable of classifying fraud transactions from legitimate ones efficiently. However, this endeavor is fraught with challenges, chief among them being the class imbalance problem and the dynamic nature of transaction data.

The issue of class imbalance, commonly encountered in credit card fraud detection, stems from the unequal distribution of fraudulent transactions relative to legitimate ones. This disproportionality presents a notable challenge for conventional machine learning algorithms, which often exhibit a bias towards the majority class, leading to less than optimal predictive performance. To solve the issue, various techniques have been proposed, including resampling methods such as under sampling and oversampling, ensemble methods like bag- ging and boosting, and cost-sensitive learning approaches. Among these techniques, Support Vector Machine (SVM)[3] emerges as a prominent contender.

SVM, a supervised learning technique, functions by identifying the optimal hyperplane to distinguish between various classes within the feature space. Its proficiency in delineating intricate decision boundaries renders it particularly suitable for tasks involving high-dimensional data, like credit card fraud detection. By discerning underlying patterns and correlations in the dataset, SVM enables precise classification of trans- actions into fraudulent or legitimate categories. However, its efficacy can be contingent upon the selection of appropriate kernel functions and parameter tuning, and its computational complexity may pose challenges for large-scale datasets.

In tandem with SVM, Random Forest (RF)[3] emerges as another powerful tool in the fraud detection arsenal. RF, an ensemble learning method based on decision trees, during training, it constructs several decision trees and combines their predictions to generate the ultimate output. Notably, RF exhibits robustness to over fitting, making it particularly well- suited for tasks with noisy or imbalanced datasets, such as credit card fraud detection. Moreover, its ability to provide insights into feature importance aids in understanding the underlying patterns driving fraudulent activities.

Despite their distinct methodologies, both SVM and RF share a common goal. Their complementary strengths and adaptability to different datasets offer promising avenues for developing robust and efficient fraud detection models. By leveraging the capabilities of SVM and RF, researchers and practitioners alike can bolster the security of financial transactions, safeguarding consumers and businesses in an increasingly digital landscape.

D. APPLIED MACHINE LEARNING AND ENSEMBLE LEARNING TECHNIQUES

This research explores how ma- chine learning (ML) can be used to identify fraudulent credit card transactions (CCF). The main method examined is the ”random forest” technique, which uses a combination of decision trees. Additionally, the paper introduces APATE, a method that combines random forests with network analysis for better detection.Beyond individual approaches, the study discusses common ML algorithms like Logistic Regression, Neural Networks, Decision Trees, Support Vector Machines, and Naive Bayes in the context of CCF detection. It highlights the importance of ”clustering” for improving the accuracy of these methods. Imagine you have a bunch of labeled pictures: cats, dogs, birds, etc. You feed this to a supervised learning algorithm. The algorithm learns what makes a cat picture a ’cat’, and so on. Now, when you give it an unlabeled picture, the KNN algorithm will check the ’k’ number of pictures closest to it. If, say, 3 out of 5 nearest pictures are cats, the new picture likely gets labeled as a ’cat’ too. You’ll often need to play around with the value of ’k’ to get the most accurate results.. An SVM is a type of computer algorithm used for classification tasks (sorting things into groups).In the case of fraud detection, it helps separate real transactions from fake ones. Focus on boundaries: The SVM finds the best line (or hyperplane in higher dimensions) to divide the real and fraudulent data. Support vectors: The key data points closest to this dividing line are called support vectors.

These are the most important for making the right classification. Adapting to new data: The SVM uses the dividing line and support vectors to decide whether new transactions are likely real or fraudulent. Handles tricky data: The SVM is good at finding patterns even when the difference between real and fake transactions isn't totally obvious. Better than some other methods: Research suggests it can outperform simpler classification methods like naive Bayes.

III. PROPOSED SYSTEM

A. Architecture Diagram

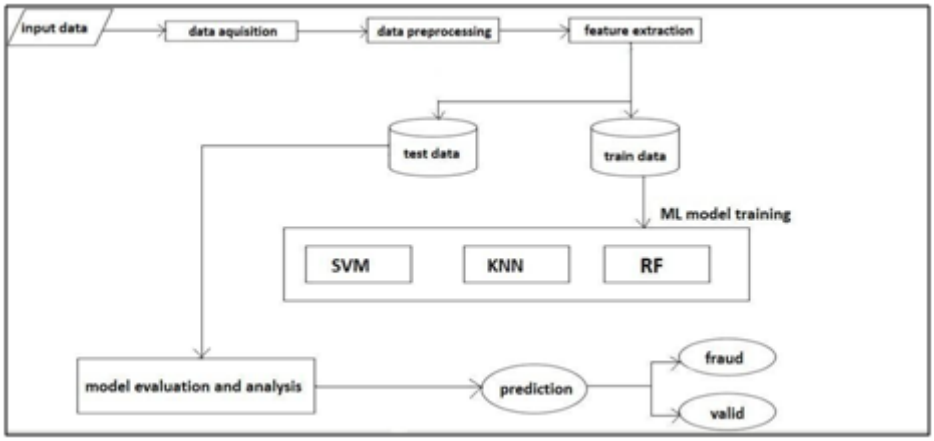


FIGURE 2: Architecture Diagram

Fig 2 is the architectural representation of the proposed system which has the main steps as Data gathering, Data pre processing, feature extraction, selecting a model, Training the model, Validation and testing , Real time monitoring etc Detection systems for credit card fraud play a vital role in recognizing and preventing unauthorized or fraud transactions. These systems leverage various techniques and technologies to analyze patterns, detect anomalies, and ensure the security of financial transactions as shown in the fig 2. Here's an outline of a proposed system:

Gathering data: Gather historical transaction details, that includes legitimate and fraud transactions. Utilize information such as transaction amount, location, time, and user behavior.

Data Pre processing: Cleaning and pre processing the collected details to handle lost values. Normalize or scale numerical features for consistent analysis. Encode categorical variables and convert timestamps into meaningful features.

Feature Engineering: taking important features that helps to classify the legitimate and the fraudulent transactions. Consider factors like transaction frequency, average transaction amount, time of day, geographical location, and user behavior.

Selecting a model: select machine learning algorithms for detecting fraud. Common algorithms considered are Logistic Regression(LR), Decision Trees(DT), Random Forests(RF), Support Vector Machines (SVM), Neural Networks(NN), Anomaly Detection Algorithms like (e.g., Isolation Forest, One-Class SVM).

Training: Partition the dataset into training and testing sub- sets to assess the model's effectiveness. Train the chosen model using the training data,fine-tuning hyper parameters as necessary. Employ methods like oversampling the minority class(fraud transactions) to mitigate class imbalance.

Validation and Testing: Assess the model's performance on the testing set for generalizability. Utilize metrics like precision, recall, f1-score,and area under the Receiver Operating Characteristic(ROC) curve to gauge the model's effectiveness during validation and testing.

Real-time Monitoring: Implement the model in a real-time system to monitor transactions as they occur. Set up alert mechanisms to notify relevant parties if a potentially fraudulent transaction is detected. Continuous **Learning:** Implement mechanisms for continuous learning to adapt the model to evolving patterns of fraud. Continuously update the model with fresh data to enhance its accuracy and efficacy over time.

Integration with Fraud Prevention Tools: Integrate the financial transaction fraud prevention system with other security measures such as two-factor authentication and card blocking mechanisms.

Compliance and Regulations: Ensure that the proposed system complies with relevant financial regulations and data protection laws.

User Education: Educate users on best practices for securing their credit card information and recognizing potential fraud.

Implementing a comprehensive financial transaction fraud detection system involves combination of advanced algorithms, data processing techniques, and ongoing monitoring to stay ahead of evolving fraud tactics. Additionally, collaboration with financial institutions and adherence to industry standards are essential for building an effective and secure system.

IV. COMPARATIVE STUDY

Table 1 presents a comparative analysis from the research paper, which explores different algorithms aimed at addressing credit card fraud detection.Through this comprehensive examination in table 1,we explore the complexities of detecting credit card fraud by thoroughly examining diverse machine learning methodologies. We rigorously evaluate the effectiveness of different strategies, including the integration of SMOTE and AdaBoost, as well as the utilization of AdaBoost in conjunction with majority voting. Our examination encompasses advanced techniques in both machine learning and deep learning aimed at improving credit card fraud detection. Furthermore, we investigate an intelligent approach, utilizing an optimized light gradient boost- ing machine. Through a detailed analysis of these methods, our goal is to ascertain their efficiency in bolstering the security of credit card transactions, offering a nuanced comprehension of their strengths and potential applications in the ongoing fight against fraudulent activities.

Paper	Advantages	Disadvantages
Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost.	--Improved detection --Robustness	--Long-training test --Require expertise
Credit card fraud detection using AdaBoost and majority voting	--Improved accuracy --Reduced overfitting	--Expensive to train --Increased computational complexity
Deep Representation Learning With Full Center Loss for Credit Card Fraud Detection	--Good and stable performance --Supervised learning from distance and angle	--Data Dependence --Loss Function Complexity
Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms	--Visualization Techniques --Focus on Popular Models	--Lack of Sources --Lack of Clarity

TABLE 1: Comparative Study of the papers

V. CONCLUSION

The study analyzed credit card fraud detection using SVM, RF, and KNN algorithms. KNN showed the highest accuracy of 99.95, indicating its effectiveness in detecting fraudulent links. While SVM and RF performed well, the 99.99 accuracy suggests improvement. The findings highlight the potential of machine learning in fraud prevention, with SVM being the most effective. Further research is needed to generalize these findings to different data types and explore cluster methods.

VI. REFERENCES

[1] E. Ileberi, Y. Sun and Z. Wang, "Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost," in IEEE Access, vol. 9, pp. 165286-165294, 2021, doi:10.1109/ACCESS.2021.3134330.

[2] Kuldeep Randhawa, Chu Kiong Loo, anjeevan Seera , Chee Peng Lim,Asoke K. Nandi, "Credit card fraud detection us- ing AdaBoost and majority voting," in IEEE Access, 2017, doi:10.1109/ACCESS.2018.2806420

[3] Z. Li, G. Liu and C. Jiang, "Deep Representation Learning With Full Center Loss for Credit Card Fraud Detection," in IEEE Transactions on Computational Social Systems, vol. 7, no. 2, pp. 569-579, April 2020, doi: 10.1109/TCSS.2020.2970805.

[4] Fawaz Khaled Alarfaj Muhammadramzan, Iqra Malik, Hikmat Ul- lah Khan,Andmuzamilahmes, " Credit Card Fraud Detection Using State of-the-Art Machine

International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

A Survey on Visually Impaired Teacher Support System

Ms. Elizabeth Anns*[1], Melvin Biju[2], Mohamed Fahad[2], Samuel Joseph[2], Sohit S[2]
*[1]Assistant Professor, Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Muvattupuzha, Kerala, India
[2]Department of Computer Science and Engineering, Viswajyothi College of Engineering and Technology, Muvattupuzha, Kerala, India

ABSTRACT
This paper delves into three influential computer vision algorithms YOLO, Faster R-CNN, and R-FCN highlighting their strengths and applications. YOLO excels in real-time action recognition, Faster R-CNN enhances behavior detection in classrooms, and R-FCN specializes in hand-raising gestures. A comparative study evaluates their advantages and disadvantages. The paper concludes with an overview of a Visually Impaired Teacher Support System leveraging YOLOv5 for enhanced classroom monitoring, providing insights for algorithm selection in computer vision applications.

Keywords: YOLO, Faster R-CNN, R-FCN, Computer Vision, Real-time Action Recognition, Behavior Detection, Hand-raising Gesture Detection, Algorithm Comparison.

I. INTRODUCTION

YOLO excelling in real-time action recognition [3], Faster R-CNN [1] enhancing behaviour detection in dynamic environments like classrooms, and R-FCN [4] demonstrating prowess in detecting intricate hand-raising gestures. The paper meticulously dissects the principles, achievements, and innovations inherent in each algorithm, providing an in-depth understanding of their individual strengths and applications. YOLO's swift video frame processing, leading to an impressive 85 percent accuracy in real-time action recognition on datasets like LIRIS Human Activities, underscores its efficacy in applications like video surveillance and human-computer interaction. However, the paper also conscientiously acknowledges the challenges, such as the speed-accuracy trade-off and ethical concerns, especially in the realm of video surveillance. Faster R-CNN, with its custom modifications and integration of ResNet-101, emerges as a linchpin in behaviour detection within classroom settings. Leveraging a feature pyramid based on a 101-layer ResNet, this algorithm excels in extracting features at different scales, facilitating the detection of small objects like hands. The adaptive template selection strategy and feature pyramid strategy introduce innovative approaches, enhancing hand-raising detection accuracy to an impressive average of 85 percent on a real classroom dataset. YOLO (You Only Look Once), Faster R-CNN (Region-Based Convolution Neural Network), and R-FCN (Region-based Fully

Convolutional Network) algorithms have demonstrated significant advancements in real-time action recognition, behaviour detection, and hand-raising gesture detection

The rest of the paper is structured as follows. Section II tells about the three algorithms YOLO, R-CNN and R-FCN and its operations. Section III discusses about the proposed system, architecture and the implementation of the Visually Impaired Teacher Support System. IV presents the comparative study on the algorithms and section V concludes the paper.

II. LITERATURE SURVEY

A. YOLO (You Only Look Once) Algorithm

The You Only Look Once (YOLO) algorithm [3] stands out for its prowess in real-time action recognition, boasting an impressive accuracy rate of 85 percent on the LIRIS Human Activities dataset. What sets YOLO apart from traditional methods is its ability to swiftly process video frames, enabling rapid identification and localization of human actions. This swift processing is particularly valuable in applications such as video surveillance and human-computer interaction, where timely action is essential. Despite its efficiency, the YOLO algorithm grapples with challenges related to the speed-accuracy trade-off. While it excels in real-time processing, achieving high accuracy levels can sometimes come at the cost of processing speed. Balancing these competing demands is crucial for optimizing the performance of YOLO-based systems across various applications. The widespread adoption of YOLO-based systems in video surveillance raises ethical concerns that demand careful consideration. The pervasive nature of surveillance technologies raises questions about privacy infringement and data protection. Responsible deployment of YOLO-based surveillance systems requires robust privacy policies and safeguards to mitigate potential risks and ensure compliance with ethical standards. In the realm of education, the YOLOv3-based system [2] introduces a novel application of real-time attention level assessment. By leveraging YOLOv3's capabilities, educators gain the ability to instantly gauge students' engagement levels during classroom sessions. This real-time assessment empowers educators to adapt their teaching approaches on-the-fly, fostering a more dynamic and responsive learning environment. The integration of real-time assessment technologies in education also raises ethical considerations, particularly concerning student privacy. Balancing the benefits of real-time assessment with the need to protect students' privacy requires careful navigation. Educators and policymakers must collaborate to establish clear guidelines and protocols for the responsible and thoughtful integration of such technologies in educational settings. This includes ensuring transparent data handling practices, obtaining informed consent from students, and implementing robust security measures to safeguard sensitive information. By addressing these ethical concerns, YOLO-based systems can play a transformative role in enhancing educational practices while upholding the principles of privacy and ethics.

B. RCNN (Region Based Convolution Neural Network)

The Faster R-CNN framework [1] serves as a cornerstone in advancing behavior detection capabilities within classroom environments. At the heart of this framework lies the Scale-Aware Detection Head, a groundbreaking feature that revolutionizes how the system adapts to the complexities of dynamic classroom settings. By incorporating multiple branches with distinct dilation rates, this innovative approach effectively addresses the challenges posed by diverse student postures, ensuring the system's ability to accurately detect

objects of varying sizes. One of the standout aspects of the Scale-Aware Detection Head is its ability to overcome obstacles such as low-resolution hand gestures and diverse student postures. This adaptability is crucial for ensuring that the system remains effective in real-world classroom scenarios, where students may exhibit a wide range of behaviors and movements. The system also leverages a Feature Fusion Strategy, which combines feature pyramids and template selection techniques to enhance the automatic detection of specific student behaviors. By strategically integrating these approaches, the system can identify hand-raising, standing, and sleeping behaviors among students with remarkable precision and accuracy. The efficacy of the Scale-Aware Detection Head and Feature Fusion Strategy is further demonstrated in the system's performance evaluation on a real-world classroom dataset. Outperforming state-of-the-art baselines, the system achieves an impressive average precision of 87.3 percent for hand-raising, standing, and sleeping behaviors. This remarkable level of accuracy underscores the system's proficiency in accurately identifying and categorizing diverse hand gestures and postures within authentic classroom environments.

C. R-FCN (Region-based Fully Convolutional Network)

In the domain of hand-raising gesture detection, an advanced R-FCN (Region-based Fully Convolutional Network) architecture has been implemented, presenting a significant leap in accuracy and precision. This architecture is built upon three core components: region proposal, position-sensitive score maps, and region classification. Leveraging a feature pyramid based on a 101-layer ResNet enables the extraction of features at multiple scales, crucial for detecting small objects such as hands within varying classroom environments. The integration of position-sensitive ROI pooling enhances localization accuracy, refining hand location through precise bounding box regression. What sets this enhanced R-FCN apart are two innovative strategies tailored specifically for hand-raising detection. Firstly, the adaptive template selection strategy employs k-means++ clustering to dynamically generate templates, departing from traditional hand-picked templates. This data-driven approach ensures that the templates accurately represent real hand sizes, leading to more precise proposal generation and ultimately enhancing hand-raising detection accuracy. The feature pyramid strategy addresses inherent challenges related to low-resolution hand gestures and diverse backgrounds. By incorporating both bottom-up and top-down pathways, this strategy adeptly captures fine-grained details while simultaneously extracting broader semantic information. This holistic approach significantly improves the system's ability to detect hand-raising gestures amidst varying classroom contexts. The proposed methodology has been rigorously evaluated on a real classroom dataset, yielding an impressive average accuracy of 85 percent. This validation underscores the efficacy of the approach in hand-raising gesture detection, reaffirming its potential to enhance classroom interaction and engagement through precise and reliable detection mechanisms.

III. PROPOSED SYSTEM

The Visually Impaired Teacher Support System is an innovative solution designed to enhance classroom monitoring and management. Utilizing a camera setup and YOLOv5 Object Detection, the system can identify and classify various objects, including students and chairs. Specialized modules address specific concerns, such as the Sleep Detection Module for identifying signs of student drowsiness and the Walkout Detection Module for recognizing students leaving the classroom. The Hand Raise Detection Module enables visually impaired

teachers to identify students seeking attention. With an Alerting System for timely intervention, the system notifies teachers through sound notifications or messages on a user-friendly interface, providing real-time monitoring with live video feeds and overlays indicating detected objects and alerts.

A. Architecture

The Visually Impaired Teacher Support System represented in figure 1 uses strategically placed cameras and YOLOv5 Object Detection to identify objects, including students and chairs. Specialized modules for Sleep Detection, Walkout Detection, and Hand Raise Detection, along with the Alerting System, ensure timely communication with visually impaired teachers. The user-friendly interface provides live video feeds and overlays for effective classroom management, enabling quick interventions in inclusive environments.

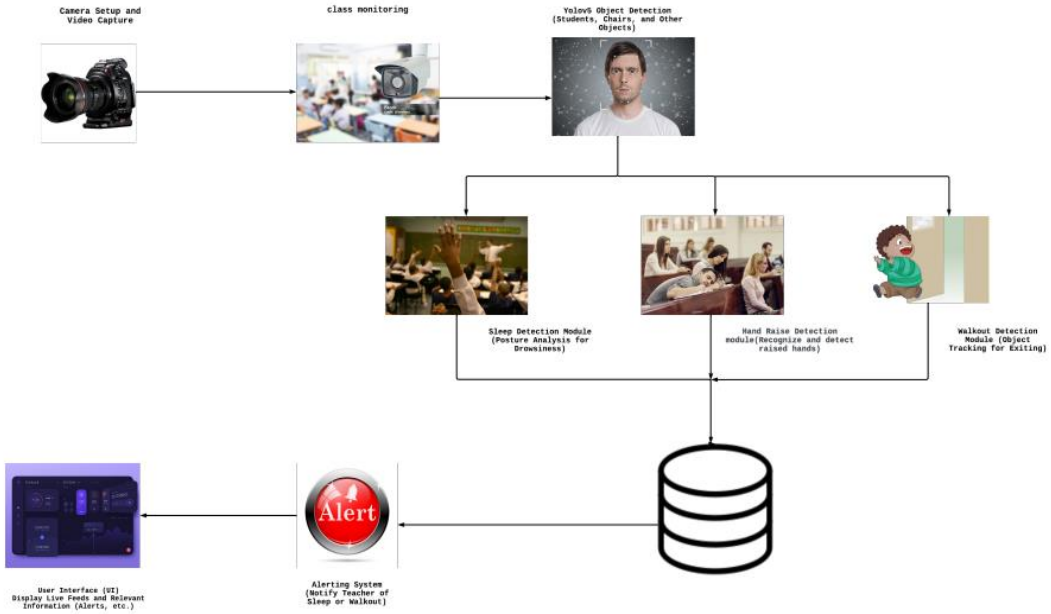


Figure 1: Architecture Diagram

B. Implementation

- Video Capture: Cameras capture real-time footage for subsequent analyses, ensuring comprehensive coverage.
- YOLOv5 Object Detection: Utilizes cutting-edge YOLOv5 algorithm for real-time identification and classification of objects, enhancing understanding of the classroom environment.
- Sleep Detection Module: Employs computer vision techniques to analyze student body postures, triggering alerts for potential drowsiness.
- Walkout Detection Module: Utilizes object tracking algorithms to monitor student movement, generating alerts when a student leaves the frame.
- Hand Raise Detection Module: Dedicated module identifies instances of students raising hands, facilitating timely teacher responses.
- Alerting System: Ensures timely communication with visually impaired teachers through sound notifications or messages.

- User Interface (UI): Crafted for accessibility, provides a user-friendly platform with live video feeds and overlays for detailed classroom monitoring.
- Comprehensive Monitoring: Addresses various classroom dynamics, offering a holistic solution for effective management.
- Timely Intervention: Enables prompt responses to student activities, signs of drowsiness, or students leaving the classroom, contributing to controlled teaching environments.

IV. COMPARATIVE STUDY

In table 1 the three key algorithms YOLO, Faster R-CNN, and R-FCN are critically analyzed to evaluate their strengths, limitations, and applications in real-time action recognition, behavior detection, and hand-raising gesture detection. This concise examination aims to offer insights for researchers and practitioners, aiding in the informed selection of algorithms based on specific application requirements. The section acts as a valuable resource, guiding decision-making in the dynamic realm of computer vision and pattern recognition.

Paper	Advantage	Disadvantage
Intelligent Student Behaviour Analysis System for Real Classrooms	Automation Accuracy	Data Dependency Real Time Challenges
Classroom behaviour recognition based on improved yolov3	Adaptability Speed	Accuracy Resource Intensive
Hand-Raising gesture detection in real classroom using improved R-FCN	High accuracy Practical use	Heavy computing Limited Generalization
Real-time Gender Identification from Face Images using You Only Look Once (YOLO)	High Accuracy Custom Dataset	Limited Accuracy for Complex Cases Dependency on High-Quality Images

Table 1: Algorithm Comparison

V. CONCLUSION

The paper explores three crucial algorithms YOLO, Faster R-CNN, and R-FCN each playing a significant role in computer vision. YOLO excels in real-time action recognition, Faster R-CNN enhances behavior detection, and R-FCN demonstrates prowess in detecting intricate hand-raising gestures. The Visually Impaired Teacher Support System integrates YOLOv5 algorithm for effective classroom monitoring. Faster R-CNN, with its innovative modifications, showcases adaptability in dynamic classroom environments, outperforming baselines with an average precision of 87.3 percent. R-FCN, focused on hand-raising gestures, achieves an impressive 85 percent accuracy. The proposed system incorporates YOLOv5 Object Detection and specialized modules for comprehensive classroom management. The comparative study aids researchers and practitioners in informed algorithm selection based on specific application requirements

VI. REFERENCES

[1] R. Zheng, F. Jiang and R. Shen, ”Intelligent Student Behavior Analysis System for Real Classrooms,” ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 2020, pp. 9244-9248, doi: 10.1109/ICASSP40776.2020.9053457.]

[2] Y. Zhang et al., ”Classroom behavior recognition based on improved yolov3,” 2020 International Conference on Artificial Intelligence and Education (ICAIE), Tianjin, China, 2020, pp. 93-97, doi:10.1109/ICAIE50891.2020.00029

[3] V. E.K. and C. Ramachandran, ”Real-time Gender Identification from Face Images using you only look once (yolo),” 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), Tirunelveli, India, 2020, pp. 1074-1077, doi: 10.1109/ICOEI48184.2020.9142989.

[4] Si, Jiaxin Lin, Jiaojiao Jiang, Fei Shen, Ruimin. (2019). Hand-raising Gesture Detection in Real Classrooms Using Improved R-FCN. Neurocomputing.359.10.1016/j.neucom.2019.05.031

1st International Conference on Security, Parallel Processing, Image Processing and Networking [SPIN-2K24]

Organized By

Department of Computer Science & Engineering,
Viswajyothi College of Engineering and Technology,
Muvattupuzha, Ernakulam, Kerala, India
In Association With
CSI, ISTE, R&D & A2Z Edulearning Hub

Publisher

Technoscience Academy

Website : www.technoscienceacademy.com

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