© 2019 IJSRST | Volume 6 | Issue 3 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology
DOI : https://doi.org/10.32628/IJSRST19633

Disease Prediction by Machine Learning from Healthcare Communities

Saiesh Jadhav[1], Rohan Kasar[2], Nagraj Lade[3], Megha Patil[4], Prof. Shital Kolte[5]

[1-4]Department of Computer Engineering, SPPU, Pune, Maharashtra, India

[5]Assistant Professor, Department of Computer Engineering, SPPU, Pune, Maharashtra, India

ABSTRACT

To promote sustainable improvement, the smart town implies a global imaginative and prescient that merges artificial intelligence, choice making, statistics and conversation era (ICT), and the net-of-things (IoT). in this mission, the subject of disease prediction and prognosis in clever healthcare is reviewed. due to records progress in biomedical and healthcare groups, correct have a look at of clinical data advantages early disorder recognition, patient care and network services. whilst the exceptional of medical information is incomplete the exactness of study is reduced. moreover, exclusive areas exhibit specific appearances of certain regional illnesses, which can also bring about weakening the prediction of sickness outbreaks. within the proposed system, it offers gadget gaining knowledge of algorithms for effective prediction of various disorder occurrences in ailment-frequent societies and predicts the waiting time for each treatment project for every patient as well as a hospital Queuing advice (HQR) system is advanced for recommending treatment mission sequence with appreciate to anticipated ready time. It experiments on a nearby chronic illness of cerebral infarction. using structured and unstructured facts from health centre it makes use of system studying selection Tree algorithm and KNN algorithm. To the first-rate of our knowledge inside the place of medical huge records analytics none of the existing paintings focused on each information types. in comparison to several normal estimate algorithms, the calculation exactness of our proposed set of rules reaches 94.8% with a convergence speed which is faster than that of the CNN-based totally uni-modal ailment threat prediction (CNN-UDRP) algorithm. similarly, challenges within the deployment of sickness diagnosis in healthcare had been mentioned.

Keywords : Data analytics; Machine Learning; Healthcare, K Nearest Neighbor, (KNN) and Support Vector Machine (SVM), electronic health records (EHR).

I. INTRODUCTION

Machine learning is a field of software engineering that frequently utilizes factual procedures to enable PCs to "learn" (i.e., continuously enhance execution on a particular assignment) with information, without being expressly programmed. Machine learning is firmly identified with (and regularly covers with) computational measurements, which additionally centers around expectation making using PCs. It has solid connections to scientific enhancement, which conveys techniques, hypothesis and application areas to the field. Machine learning is in some cases conflated with information mining, where the last sub-field concentrates more on exploratory information investigation and is known as unsupervised learning. Machine learning can likewise be unsupervised and be utilized to learn and set up benchmark conduct profiles for different entities and after that used to discover significant abnormalities. Inside the field of data analytics, machine learning is a technique used to devise

complex models and calculations that loan themselves to forecast; in business utilize, this is known as prescient examination. These scientific models permit specialists, information researchers, architects, and investigators to "deliver dependable, repeatable choices and comes about" and reveal "concealed bits of knowledge" through gaining from authentic connections and patterns in the data.

Currently, most hospitals are overcrowded and lack effective patient queue management. Patient queue management and wait time prediction form a challenging and complicated job because each patient might require different phases/ operations, such as a check-up, various tests, e.g., a sugar level or blood test, X-rays or a CT scan, minor surgeries, during treatment. We call each of these phases /operations as treatment tasks or tasks in this paper. Each treatment task can have varying time requirements for each patient, which makes time prediction and recommendation highly complicated. A patient is usually required to undergo examinations, inspections or tests (refereed as tasks) according to his condition. In such a case, more than one task might be required for each patient. Some of the tasks are independent, whereas others might have to wait for the completion of dependent tasks. Most patients must wait for unpredictable but long periods in queues, waiting for their turn to accomplish each treatment task. In this paper, we focus on helping patients complete their treatment tasks in a predictable time and helping hospitals schedule each treatment task queue and avoid overcrowded and ineffective queues. We use massive realistic data from various hospitals to develop a patient treatment time consumption model. The realistic patient data are analysed carefully and rigorously based on important parameters, such as patient treatment start time, end time, patient age, and detail treatment content for each different task. We identify and calculate different waiting times for different patients based on their conditions and operations performed during treatment.

II. LITERATURE REVIEW

2.1 Jianguo Chen, Kenli Li, Zhuo Tang, Kashif Bilal, And Keqin Li “A Parallel Patient Treatment Time Prediction Algorithm and Its Applications in Hospital Queuing-Recommendation in A Big Data Environment” [1]. Author propose a Patient Treatment Time Prediction (PTTP) algorithmic to predict the waiting time for every treatment task for a patient. Supported the expected waiting time, a Hospital Queuing Recommendation (HQR) system is developed. HQR calculates Associate in Nursing predicts an economical and convenient treatment set up suggested for the patient. As a result of the large-scale, realistic data-set and also the demand for time period response, the PTTP algorithmic and HQR system mandate potency and low-latency response. Our proposed model to recommend an effective treatment plan for patients to minimize their wait times in hospitals.

2.2 L. Guo, C. Zhang, J. Sun, Y. Fang. “A privacy-preserving attribute-based authentication System for Mobile Health Networks,” [2]. Electronic healthcare (eHealth) systems have replaced paper-based medical systems due to the attractive features such as universal accessibility, high accuracy and low cost. As a major component of eHealth systems, mobile healthcare (mHealth) applies mobile devices, such as smartphones and tablets, to enable patient-to-physician and patient-to-patient communications for better healthcare and quality of life (QoL). Unfortunately, patients’ concerns on potential leakage of personal health records (PHRs) are the biggest stumbling block. In current eHealth/mHealth networks, patients’ medical records are usually associated with a set of

attributes like existing symptoms and undergoing treatments based on the information collected from portable devices. To guarantee the authenticity of those attributes, PHRs should be verifiable. However, due to the linkability between identities and PHRs, existing mHealth systems fail to preserve patient identity privacy while providing medical services. To solve this problem, we propose a decentralized system that leverages users' verifiable attributes to authenticate each other while preserving attribute and identity privacy.

2.3 A. Abbas, S. Khan, "A review on the state-of-the-art privacy preserving approaches in e-health clouds," [3]. Cloud computing is emerging as a new computing paradigm in the healthcare sector besides other business domains. Large numbers of health organizations have started shifting the electronic health information to the cloud environment. Introducing the cloud services in the health sector not only facilitates the exchange of electronic medical records among the hospitals and clinics, but also enables the cloud to act as a medical record storage center. This survey aims to encompass the state-of-the-art privacy preserving approaches employed in the e-Health clouds. Moreover, the privacy preserving approaches are classified into cryptographic and non-cryptographic approaches and taxonomy of the approaches is also presented. Furthermore, the strengths and weaknesses of the presented approaches are reported and some open issues are highlighted.

2.4 J. Yang, J. Li, Y. Niu, "A hybrid solution for privacy preserving medical data sharing in the cloud environment," [4].
Storing and sharing of medical data in the cloud environment, where computing resources including storage is provided by a third-party service provider, raise serious concern of individual privacy for the adoption of cloud computing technologies. This paper proposes a practical solution for privacy preserving medical record sharing for cloud computing. Based on the classification of the attributes of medical records, we use vertical partition of medical dataset to achieve the consideration of different parts of medical data with different privacy concerns.

2.5 V. Goyal, O. Pandey, A. Sahai, B. Waters, "Attribute-based encryption for fine-grained access control of encrypted data," [5]. Authors develop a new cryptosystem for fine-grained sharing of encrypted data that we call Key-Policy Attribute-Based Encryption (KPABE). In our cryptosystem, cipher texts are labelled with sets of attributes and private keys are associated with access structures that control which cipher texts a user is able to decrypt. We demonstrate the applicability of our construction to sharing of audit-log information and broadcast encryption. Our construction supports delegation of private keys which subsumes Hierarchical Identity-Based Encryption (HIBE).

2.6 J. Han, W. Susilo, Y. Mu. "Improving privacy and security in decentralized cipher text-policy attribute-based encryption," [6].
In previous privacy-preserving multi-authority attribute-based encryption (PPMA-ABE) schemes, a user can acquire secret keys from multiple authorities with them knowing his/her attributes and furthermore, a central authority is required. Notably, a user's identity information can be extracted from his/her some sensitive attributes. Hence, existing PPMAABE schemes cannot fully protect users' privacy as multiple authorities can collaborate to identify a user by collecting and analyzing his attributes. Moreover, cipher text-policy ABE (CPABE) is a more efficient public-key encryption where the encrypt or can select flexible

access structures to encrypt messages. Therefore, a challenging and important work is to construct a PPMA-ABE scheme where there is no necessity of having the central authority and furthermore, both the identifiers and the attributes can be protected to be known by the authorities. In this paper, a privacy-preserving decentralized CP-ABE (PPDCPABE) is proposed to reduce the trust on the central authority and protect users' privacy.

III. PROBLEM STATEMENT

Current scheme has some defects. like the records set is usually small, for sufferers and illnesses with specific situations, the characteristics are selected through experience. But those pre-selected traits perhaps not fulfil the adjustments inside the disease and its influencing elements. To conquer these issues, on this paper, we advise a brand new convolutional neural network based totally multimodal disease danger prediction (CNN-MDRP) algorithm the usage of established and unstructured information from medical institution. The prescribed tests ready time of each remedy undertaking is acquired by means of them.

PTTP model, which is the sum of all sufferers' probably treatment times in the contemporary queue. An HQR machine is proposed primarily based at the predicted waiting time. A remedy recommendation with a green and handy treatment plan and the least waiting time is recommended for every affected person. The PTTP algorithm and HQR device are sizable sanatorium information are saved inside the database.

IV. SYSTEM ARCHITECTURE OVERVIEW

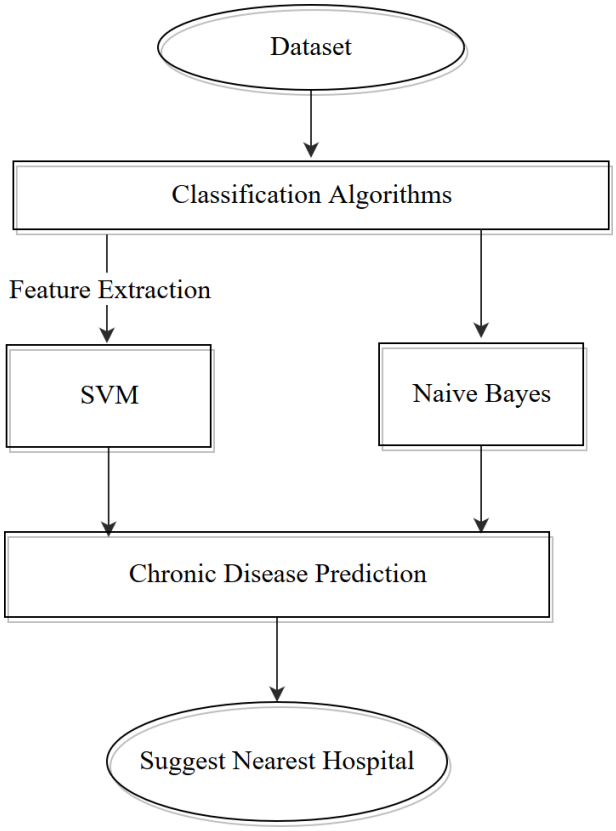

Fig.1 System Architecture

In the proposed gadget, it presents gadget getting to know algorithms for powerful prediction of numerous ailment occurrences in ailment-common societies and predicts the waiting time for each treatment challenge for every patient as well as a health centre Queuing advice (HQR) machine is advanced for recommending treatment challenge series with admire to predicted waiting time. It experiments on a local persistent illness of cerebral infarction. The use of dependent and unstructured data from health centre it makes use of system getting to know selection Tree algorithm and KNN set of rules. To the satisfactory of our expertise in the location of medical large information analytics none of the present work targeted on both information types. In comparison to numerous common estimate algorithms, the calculation exactness of our proposed set of rules reaches 94.8% with a convergence pace that is quicker than that of

the CNN-based uni-modal sickness chance prediction (CNN-UDRP) algorithm. further, challenges inside the deployment of disorder analysis in healthcare were mentioned.

V. SYSTEM ANALYSIS

The aim of this paper is to are expecting whether or not an affected person is among the cerebral infarction excessive-danger population consistent with their clinical records. Extra formally, we regard the chance prediction version for cerebral infarction because the supervised studying methods of machine learning, i.e., the input price is the attribute cost of the affected person, which incorporates the affected person's personal information together with age, gender, the superiority of signs and symptoms, and residing behavior (smoking or no longer) and other structured information and unstructured records. The output price indicates whether or not the patient is amongst the cerebral infarction excessive-risk population.

In keeping with the distinct characteristics of the patient and the discussion with docs, we will consciousness on the subsequent 3 datasets to reach an end.

Dependent statistics (S-facts): use the patient's dependent records to expect whether or not the affected person is at high-threat of cerebral infarction.

Text facts (T-facts): use the affected person's unstructured text information to predict whether or not the patient is at high-chance of cerebral infarction.

Established and textual content information (S&T-statistics): use the S-facts and T-data above to multi-dimensionally fuse the structured records and unstructured text information to predict whether or not the affected person is at excessive-danger of cerebral infarction.

On this paper, for S-statistics, in step with the dialogue with medical doctors and Pearson's correlation analysis, we extract the affected person's demographics characteristics and some of the characteristics related to cerebral infarction and dwelling conduct (which include smoking).

VI. CONCLUSION

Information mining helps many one of the kind strategies for knowledge discovery and prediction collectively with elegance, clustering, sequential pattern mining, affiliation rule mining and analysis. information mining is significantly applied in enterprise evaluation, strategic choice making, economic forecasting; destiny sales prediction and so forth. Device studying algorithms are proposed for effective prediction of continual sickness.

To extract feature from unstructured facts RNN algorithm can be used. Right here, character will upload the take a look at report i.e. preceding fitness record. RNN set of rules extracts the abilities from that file and passes that capabilities to the Naïve SVM algorithm for ailment predication.
Device proposes Naive Bayesian set of rules to predict the sickness the usage of primarily based facts. System lets in person to pick the symptoms. machine passes those symptoms to the Naive Bayes algorithm to perform disease prediction.

In the proposed device, it affords system getting to know algorithms for effective prediction of

numerous disorder occurrences in disease-common societies and predicts the waiting time for each remedy task for each affected person in addition to a Hospital Queuing Recommendations (HQR) gadget is evolved for recommending remedy mission series with appreciate to anticipated waiting time.

VII.ACKNOWLEDEGMENT

I would prefer to give thanks the researchers likewise publishers for creating their resources available. I'm conjointly grateful to guide, reviewer for their valuable suggestions and also thank the college authorities for providing the required infrastructure and support.

VIII. REFERENCES

[1]. Min Chen, Yixue Hao, Kai Hwang, Lu Wang, and Lin Wang, "Disease Prediction by Machine Learning over Big Data from Healthcare Communities", IEEE transaction, 2017, pp 8869-8879.

[2]. W. Yin and H. Schutze," Convolutional neural network for paraphrase identification", in HLTNAACL,2015, pp. 901-911.

[3]. Seema sharma, Jitendra Agarwal, Shikha Agarwal, Sanjeev Sharma, "Machine Learning Techniques for Data Mining: A Survey", in Computational Intelligence and Computing Research, IEEE International Conference on. IEEE, 2013, pp.1-6.

[4]. Jensen PB, Jensen LJ, Brunak S, "Mining electronic health records: towards better research applications and clinical care," Nat Rev Genet.2013 Jan; 14(1):75.

[5]. L. Qiu, K. Gai, and M. Qiu, "Optimal big data sharing approach for tele-health in cloud computing", in Smart Cloud (Smart Cloud), IEEE International Conference on. IEEE, 2016, pp. 184-189.

[6]. Siwei Lai, Xu Kang Liu, Jun Zhao," Recurrent Convolutional Neural Networks for Text Classification", in proceeding of the twenty-ninth AAAI Conference on Artificial Intelligence 2015.

[7]. Xingyou Wang, Weijie Jiang, Zhiyong Luo, "Combination of Convolutional and Recurrent Neural Network for Sentimental Analysis of Short Texts", International Conference on Computational Linguistics: technical papers, 2016, pg 2428-2437

[8]. Dipak V.Patil, R.S. Bichkar, "Multiple Imputation of Missing Data with Genetic Algorithm based Techniques", IJCA Special issue on Evolutionary Computation for Optimization Technique, 2010.

[9]. Ying Wen, Weinan Zhang, Rui Luo, Jun Wang, "Learning text representation using recurrent convolutional neural network with highway letters", Neu-IR 16 SIGIR Workshop on Neural Information Retrieval, July 21,2016, Pisa, Italy.

[10]. N. Nori, H. Kashima, K. Yamashita, H. Ikai, and Y. Imanaka, "Simultaneous modelling of multiple diseases for mortality prediction in acute hospital care", in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015.

[11]. J. C. Ho, C. H. Lee, and J. Ghosh, "Septic shock prediction for patients with missing data", ACM Transactions on Management Information Systems (TMIS), vol. 5, no. 1, p. 1, 2014.

[12]. Liqiang Nie, Xiaochi Wei, Dongxiang Zhang, Xiang Wang, Zhipeng Gao, and Yi Yang," "Data-driven Answer Selection in Community QA Systems", IEEE TRANSACTIONS ON

KNOWLEDGE AND DATA ENGINEERING JUNE 2016.

[13]. Kanchan M. Tarwani, Swathi Edem, "Survey on Recurrent Neural Network in Natural Language Processing", International Journal of Engineering Trends and Technology (IJETT), Volume 48 Number 6, June 2017.

[14]. Jan Koutnik, Klaus Greff, Faustino Gomez, Jurgen Schmidhuber, "A Clockwork RNN", IDSIA, USI SUPSI, Manno Lugano, CH-6928, Switzerland.

[15]. R. Vijaya Kumar Reddy, K. Prudvi Raju, M. Jogendra Kumar, CH. Sujatha, P. Ravi Prakash," Prediction of Heart Disease Using Decision Tree Approach", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 3, March 2016.

[16]. Networks Dimitrios H. Mantzaris, George C. Anastassopoulos and Dimitrios K. Lymberopoulos," Medical Disease Prediction Using Artificial Neural", Bio-Informatics and BioEngineering, 2008. BIBE 2008. 8th IEEE International Conference on 08 December 2008.

[17]. Youn-Jung Son, Hong-Gee Kim, Eung-Hee Kim, Sangsup Choi, and Soo-Kyoung Lee, "Application of Support Vector Machine for Prediction of Medication Adherence in Heart Failure Patients", Health Information Research, v.16, 2010.

[18]. Alaa Elsayad and Mahmoud Fakhr, "Diagnosis of Cardiovascular Diseases with Bayesian Classifiers", Department of Computers and Systems, Electronics Research Institute, 12622 Bohoth St., Dokki, Geza, Egypt, 2015.

[19]. M. Akhil jabbar, Dr.Priti Chandra, Dr. B. L Deekshatulu, "Heart Disease Prediction System using Associative Classification and Genetic Algorithm", International Conference on Emerging Trends in Electrical, Electronics and Communication Technologies,-ICECIT, 2012.

Cite this article as :

Saiesh Jadhav, Rohan Kasar, Nagraj Lade, Megha Patil, Prof. Shital Kolte, "Disease Prediction by Machine Learning from Healthcare Communities", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 6 Issue 3, pp. 29-35, May-June 2019. Available at doi : https://doi.org/10.32628/IJSRST19633
Journal URL : http://ijsrst.com/IJSRST19633