ISSN: 2581-3404 (Online)
International Journal of Innovative Research in Technology and Management, Vol-4, Issue-4, 2020.

Diseases Prediction and Analysis for Healthcare Communities: Survey and Discussions

Sanjana Chakraborty[1], Dr. Sunil Phulre[2]

[1]M. Tech. Scholar, Department of CS, LNCTU, Bhopal (India)

[2]Associate Professor, Department of CS, LNCTU, Bhopal (India)

ABSTRACT
Disease prediction has the potential to benefit stakeholders such as the government and health insurance companies. It can identify patients at risk of disease or health conditions. Clinicians can then take appropriate measures to avoid or minimize the risk and in turn, improve quality of care and avoid potential hospital admissions. Due to the digitalization of data in each and every domain it is becoming tedious to store and analysis. So, the demand of proficient algorithms for health care data analysis is also increasing. Machine learning is a subfield of Artificial Intelligence that allows a computer system to learn from the environment, through re-iterative processes and improve itself from experience. Here we present the review work for the diseases detection using the data mining and machine learning techniques.

Keywords:- Health Care, Data Mining, Machine Learning, data Warehouse, Information technology.

INRODUCTION
The World Health Organization reported that non communicable diseases contribute approximately 41 million premature deaths annually, i.e., nearly 71% of all deaths globally. If unmitigated, the total number of non communicable disease related deaths is estimated to reach 52 million yearly by 2030.

The most common non communicable diseases are diabetes and hypertension, which account for nearly 46.2% and 4% of total deaths, respectively. Type 2 diabetes is a continuing metabolic disorder that changes blood glucose levels and is usually a consequence of the body's ineffectiveness to employ its produced insulin; Individuals that have diabetes are likely to have higher risk for stroke and mortality [13].

A huge amount of data is generated by healthcare industry [6]. Healthcare industries are discovering novel analytic models to identify high risk patients, deploy evidence based medicine, and reduce adverse events and infections. New pay for-performance models will require healthcare organizations to improve quality and reduce costs by proactively classifying at-risk members of the community, identifying gaps in care, and improving patient engagement. New methodologies are needed to tailor treatment plans that risk-stratify patients, reduce the need for emergency treatments, and prevent hospital readmissions. Real-time analytics is needed to guide the way to informed diagnosis and improved outcomes. Analytics can also provide insight to optimize staffing schedules, better manage IT resources, and identify trends in services that can be used for marketing. With the large amount of data potentially available for analysis, managing the flow of data efficiently can be a challenge. Rapidly feeding the results of data analysis to multiple devices is critical for a solution.

Hence, new models are to be increased in the process of diagnosis. It should be a process, which recognizes patterns as well as the differences and closest relationship in the data. The database size employed in the data analysis was increasing at exponential rates gradually. The increasing number of the data analysis process in an organization is facing the difficulty of size of databases, and the data explosion employed in the organization has been increasing at an exponential rate.

The raw source data have been produced over a lot of sources like transactions, social networking sites, Web servers, business processes, etc., and remains in the form of controlled and the unstructured form. The big data, which are huge volume of datasets, are difficult for the method of storage and processing and analyzed by the newly available tools and database management systems. The use of big data can be continuously increased in a certain period in various fields, mainly in the purpose of diagnosing and predicting the disease of the patients [15].

II ELECTRONIC HEALTH RECORD

Health data are the single largest source of utilization data for different healthcare services. These data are generated when a healthcare consumer interacts with healthcare entities, such as visiting a general physician, having diagnostic procedures performed, being admitted to a hospital, or buying medicines at a pharmacy. The term ‘electronic health data’ can also refer to ‘administrative healthcare data’, ‘administrative claim data’, ‘healthcare utilization data’, ‘electronic health records’, ‘computerised claim data’ or ‘digitised health records’. These data are used for the care of patients, as they may contain information about diagnoses, treatments, providers and patient transfers within different functional units of a single hospital or across different healthcare facilities, such as hospitals, rehab centres and nursing homes.

In recent times, electronic health data have widely been used in the healthcare research community for research and clinical decision making, such as diagnosis, treatment and disease prediction. Traditionally, clinical decisions have depended on doctors’ intuition, knowledge and experience, with the aid of various clinical and diagnosis tests. This practice often leads to unwanted biases, errors and high expenses and negatively affects the quality of service provided to patients [8].

With the recent advent of data analysis methods and tools (e.g., big data analysis methods), electronic health data are being utilized in a wide range of healthcare research areas rather than just for record keeping, such as the analysis of healthcare utilization, measuring performance of hospital care networks [8], exploring patterns and cost of care, developing disease risk prediction models, chronic disease surveillance, and comparing disease prevalence and drug outcomes. Besides, these data are an important resource for the research and surveillance of chronic diseases. For example, highly prevalent and serious health conditions (e.g., heart failure) are often diagnosed at a later stage in their progression. If prediction models could identify these health risks earlier, patients could be treated properly, potentially changing the course of the disease or health condition.

III RELATED WORK

[1] Here they propose a new convolutional neural network based multimodal disease risk prediction (CNN-MDRP) algorithm using structured and unstructured data from hospital. To the best of our knowledge, none of the existing work focused on both data types in the area of medical big data analytics. Compared to several typical prediction algorithms, the prediction accuracy of their proposed algorithm reaches 94.8% with a convergence speed which is faster than that of the CNN-based unimodal disease risk prediction (CNN-UDRP) algorithm.[2] This paper reviews applications of machine learning (ML) predictive models in the diagnosis of chronic diseases. Chronic diseases (CDs) are responsible for a major

portion of global health costs. Patients who suffer from these diseases need lifelong treatment. Nowadays, predictive models are frequently applied in the diagnosis and forecasting of these diseases. In this study, we reviewed the state-of-the-art approaches that encompass ML models in the primary diagnosis of CD. This analysis covers 453 papers published between 2015 and 2019, and our document search was conducted from PubMed (Medline), and Cumulative Index to Nursing and Allied Health Literature (CINAHL) libraries. Ultimately, 22 studies were selected to present all modeling methods in a precise way that explains CD diagnosis and usage models of individual pathologies with associated strengths and limitations. [3] In the survey discuss Disease predicts the hospital data by using the different data mining technique. This analyse the medical data in multiple ways, like that, multidimensional ways and view based collects that data and it escapes the hard risks then, prediction is easily completed. The hospital data is classified in to two types namely, (i) structured data, (ii) Unstructured data. The concept fulfils the existing system focused both types of data prediction in medical area, that is big data analytics. There are numerous researches from various domains are continuously working towards developing Achieving Disease Prediction. The aim of this survey was to Summarize the recent researches and its demerits towards achieve Disease Prediction. [4] In this paper author discuss different aspects related to machine learning used for heart disease prediction. It throws light into methods that improve the classification performance as well. Such methods are known as feature selection methods. With such methods, the performance of ML algorithms is boosted. There are feature optimization methods as well as discussed in this paper. With all these methods, this paper provides useful insights to academia and industry with regard to heart disease prediction research. In the literature, different algorithms are found suitable for prediction of heart disease. However, they are data-driven approaches. Feature extraction, feature selection and feature optimization are important for improving classification algorithms. Classification algorithms are able to perform prediction task based on the training provided to them. Hence they are known as supervised machine learning algorithms. [6] This paper explains how healthcare organizations can build and develop their data analytics infrastructure, data science skills, and data governance processes necessary for a high-performing data lake. The huge data can be stored efficiently using Azure Data Lake and data analytics can be carried out to fascinate learning patterns which can be used to develop a support system to help the medical practitioner to detect the chances of heart diseases at an early stage and take suitable decisions accordingly. [7] This work compares the existing classification, clustering and prediction models and identifies the scope for improvements in this domain. As compared to the other methods, this research acknowledged that the improvements on accuracy can be achieved by using machine learning approaches. Other side, neural network models performing better in handling disease predictions. But, due to the complex activation methods in neural networks lead to higher time complexity. Thus, this work propose another novel predictive optimization technique in order to maintain efficiency during accuracy and time complexity trade off. The proposed OMLR algorithm exhibits 93.2% accuracy with lesser time complexity than the other models. [8] The goal of this comprehensive literature review was to discuss different risk prediction models that have been proposed based on electronic health data. Search terms were designed to find relevant research articles that utilised electronic health data to predict disease risks. Online scholarly databases were searched to retrieve results, which were then reviewed and compared in terms of the method used, disease type and prediction accuracy. This paper provides a comprehensive review of the use of electronic health data for risk prediction models. A comparison of the results from different techniques for three frequently modelled diseases using electronic health data was also discussed in this study. In addition, the advantages and disadvantages of different risk prediction models, as well as their performance, were presented.

Electronic health data have been widely used for disease prediction.

IV PROBLEM IDENTIFICATION

The valuable data about the patients and medications is digitized and saved as Electronic Health Records (EHR). EHR on a large scale allows the researchers to identify the possibilities to move the healthcare organizations towards the personalized healthcare. EHR comprises of structured data and the semi and unstructured data also. Only about one fifth of data is in a structured format which can be easily used by data scientist; but semi and unstructured production rate is many times greater than the structured one. However, to find the valuable insights, health IT research needs to process all kind of structured, semi and unstructured data. Hence, an enhanced information administration framework would help the data scientist to provide the custom-made prescription. The current IT infrastructure provides many systems such as Enterprise Data Warehouse (EDW); but it cannot address the issues related to scalability [6].

V CONCLUSION AND FUTURE SCOPE

Machine learning in healthcare services assists with investigating a large number of various information focus and recommend results, give auspicious risk scores, exact asset allotment, and has numerous different applications. The attention on the best way to utilize ML to expand persistent consideration. In this paper we present the comparative review for health care diseases diagnosis using the different machine learning techniques. In future we focus on specially heart diseases with machine learning techniques and improve the prediction results with accuracy.

REFERENCES:

[1] MIN CHEN, YIXUE HAO, KAI HWANG, LU WANG, AND LIN WANG, "Disease Prediction by Machine Learning Over Big Data From Healthcare Communities" SPECIAL SECTION ON HEALTHCARE BIG DATA, Vol-5, IEEE, 2017. Pp 8869-8879.

[2] Gopi Battineni, Getu Gamo Sagaro, Nalini Chinatalapudi, Francesco Amenta, "Applications of Machine Learning Predictive Models in the Chronic Disease Diagnosis", J. Pers. Med. 2020, pp 1-11.

[3] Shivaganga S P, Hemashree H C, "DISEASE PREDICTION AND ANALYSIS FOR HEALTHCARE COMMUNITIES", International Journal of Advances in Engineering & Technology, June, 2019, pp 45-51.

[4] Pulugu Dileep, Kunjam Nageswara Rao, Prajna Bodapati, " A Review on Machine Learning Techniques for Data-Driven Heart Disease Prediction", 2019, pp 304-316.

[5] P. Sunanda, " MACHINE LEARNING FOR DISEASE PREDICTION BY USING NEURAL NETWORKS", IJRAR May 2019, pp 712-719.

[6] Ekta Maini, Bondu Venkateswarlu, Arbind Gupta, "Data Lake-An Optimum Solution for Storage and Analytics of Big Data in Cardiovascular Disease Prediction System", International Journal of Computational Engineering & Management, 2018, pp 33-39.

[7] N Satyanandam, Ch Satyanarayana, "Heart Disease Detection Using Predictive Optimization Techniques", I.J. Image, Graphics and Signal Processing, 2019, pp 18-24.

[8] Md Ekramul Hossain, Arif Khan, Mohammad Ali Moni, Shahadat Uddin, "Use of electronic health data for disease prediction: A comprehensive literature review", IEEE, 2019, pp 1-20.

[9] Jenna Wiens , Suchi Saria, Mark Sendak, Marzyeh Ghassemi, Vincent X. Liu Finale Doshi-Velez, Kenneth Jung, Katherine Heller, David Kale, Mohammed Saeed, Pilar N. Ossorio, Sonoo Thadaney-Israni, Anna Goldenberg, "Do no harm: a roadmap for responsible machine learning for health care", Nature Medicine, 2019, pp 1-4.

[10] Latha R, Vetrivelan P, "Blood Viscosity based Heart Disease Risk Prediction Model in Edge/Fog Computing", 2019, pp 833-837.

[11] Stephen Dankwa, Wenfeng Zheng, "Special Issue on Using Machine Learning Algorithms in the Prediction of Kyphosis Disease: A Comparative Study", Journal of Appl. Sci. 2019, pp 1-32.

[12] A. SUHASINI, A.Subitha, S.Midhun, "Investigation Study on Disease Prediction Techniques with Big Data", Studies in Indian Place Names, 2020, pp 99-106.

[13] NORMA LATIF FITRIYANI, MUHAMMAD SYAFRUDIN, GANJAR ALFIAN, JONGTAE RHEE, "Development of Disease Prediction Model Based on Ensemble Learning Approach for Diabetes and Hypertension", SPECIAL SECTION ON DATA-ENABLED INTELLIGENCE FOR DIGITAL HEALTH, IEEE 2019, pp 144777-144790.

[14] Thanh Hai Nguyen, Edi Prifti, Nataliya Sokolovska, Jean-Daniel Zucker, "Disease Prediction using Synthetic Image Representations of Metagenomic data and Convolutional Neural Networks", IEEE, 2019, pp 1-7.

[15] R. Ramani, K. Vimala Devi, K. Ruba Soundar, "MapReduce-based big data framework using modified artificial neural network classifier for diabetic chronic disease prediction", Soft Computing, 2020, pp 1-12.

[16] R. Venkatesh, C. Balasubramanian, M. Kaliappan, "Development of Big Data Predictive Analytics Model for Disease Prediction using Machine learning Technique", Journal of Medical Systems, 2019, pp 1-8.

[17] N. Sowri Raja Pillai, K.Kamurunnissa Bee, J.Kiruthika, " PREDICTION OF HEART DISEASE USING RNN ALGORITHM", International Research Journal of Engineering and Technology, 2019, pp 4452-4459.

[18] Noura E. Maghawry, Samy Ghoniem, "A Proposed Internet of Everything Framework for Disease Prediction", 2019, pp 20-27.

[19] Nishita Mehta, Anil Pandit, Meenal Kulkarni, "Elements of Healthcare Big Data Analytics", Springer 2020, pp 23-29.