International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 06 | June 2019 www.irjet.net p-ISSN: 2395-0072

Survey on Risk Estimation of Chronic Disease using Machine Learning

Vijayalakshmi C.S [1], Dr. Niharika Kumar [2]

[1]*Student, Dept. of Computer Science and Engineering, BNMIT, Karnataka, India*
[2]*Associate Professor, Dept. of Computer Science and Engineering, BNMIT, Karnataka, India*

Abstract - *Healthcare is an inevitable task to be done in human life. The health care industry contains lots of medical data, therefore machine learning algorithms are required to make decisions effectively in the prediction of heart diseases. Healthcare sectors generate massive amounts of information about patients. Health care providers can take advantage of machine learning to predict the behavior of the patients. Machine learning provides a way to find the patterns and reason about the data. Recent research has developed into uniting these techniques to provide hybrid machine learning algorithms. Predicting patients' risk of developing certain diseases is an important research topic in healthcare. In this paper, we streamline machine learning algorithms for effective prediction of chronic disease outbreak in disease-frequent communities. We experiment the modified prediction models over real-life hospital data. We propose a new convolutional neural network (CNN) based multimodal disease risk prediction algorithm using structured and unstructured data from hospital. The paper additionally describes the term Unimodal Disease Risk Prediction (UDRP) and it compares and analyze with the performance.*

Key Words: **Big data Analytics, Machine Learning, Healthcare.**

1. INTRODUCTION

Chronic diseases have been among the major concerns in medical fields since they may cause a heavy burden on healthcare resources and disturb the quality of life. Now a days the healthcare fields are making great progress with the rapid development of latest information technologies. In particular, Electronic Medical Records (EMRs) have brought significant benefits to medical applications. EMRs are sequential with large-scale, which include the records of hospital visits through the whole lifetime of people, like personal profile, diagnostic records, treatment records, etc. Heart is an important organ of all living individual, which plays an essential role of blood pumping to the rest of the organs through the blood vessels of the circulatory system. If circulation of blood in body is improper the organs like brain suffer and if heart stops working altogether and death occurs. Life is completely dependent on proper working of the heart. Prediction of heart diseases is most complicated and challenging task in the field of the medical science. Heart is one of the most common reason of death in India or other Asian countries. In 2003 approx 17.3 million people died all over globe and out of this,10 million were only due to coronary heart diseases. Along without changing lifestyle there are many such factors such as smoking, alcohol, obesity. High blood pressure, diabetes etc. which are responsible for the risk of having heart problem. The small description of previous system is, it is based and used the data mining concept of machine learning algorithm for effected prediction. This prediction is predicted the disease outbreak. This heart disease outbreak is solved in disease-frequent communities.

Heart Disease

The heart is one of the body parts that are vital for every part of the body by circulating or pumping blood to each body part. If circulation of blood in body is improper the organs like brain suffer and heart stops working altogether and death occurs. So that life ultimately depends on the heart. A properly functioning heart helps the individuals to have a healthy life. Prediction of cardiovascular disease is challenging and more complicated task to achieve an automatic diagnosis of sickness. Because an enormous amount of data are stored in healthcare centers that are very complex and challenging to analyses. Even if it is challenging task using prediction of heart diseases in medical centers is plays significant roles to save the lifestyle of individuals and to make active and accurate decision-making for stakeholders.

Some of the most common heart diseases are listed in the table below with their description

TABLE - 1: TYPES OF HEART DISEASE

Types of heart diseases	Description
Arrhythmia	Atypical heart rhythm
Acute coronary syndrome	Blood supply to the heart muscle is swiftly obstructed
Coronary artery disease	Occurred when Arteries supplying blood becomes obstructed
Angina	Chest pain due to a deficiency of blood to the heart muscle
Congenital heart disease	Heart disfigurements that are present at birth
Cardiomyopathy	Heart muscle disease
Rheumatic heart diseases	Rheumatic fever

There are also different heart disease factors, from that most common are listed in the table below with their symptom.

TABLE - 2: FACTORS OF HEART DISEASE

Risk factors	Description
Age	Old people suffers from heart disease
Sex	Males are at greater risk than females
Family history	If relatives have heart disease the probability of a person to have cardiovascular disease is high
Smoking	Heart disease higher in smokers than nonsmokers people
Poor diet	Diet food is essential for development of heart
Blood pressure	Blood pressure can effect in narrowing hardening arteries, as well as thickening blood vessels
High blood cholesterol levels	It increases formation of plaques
Diabetes	It is the disease as a result of sugar in our body
Obesity	Overweight body is one of the cause for heart diseases
Physical inactivity	Physical activity helps heart to function properly
Stress	Damage arteries
Poor hygiene	It increases heart disease

2. LITERATURE REVIEW

In [1] author has presented the concept namely, "Disease prediction using Machine Learning over Big Data". The big data is fastest concept in current trend, so this concept is applied in more fields. The big data is most widely used in each every field because it is very large. The big data is applied in medical field both side developing the better growth in both fields, that is big data is applied in medical fields develops the medical fields at the same time increase the growth in big data field. The big data helps to achieve the better growth in medical and health care sectors. It additionally, provides the more merits gives, (i) medical data analysis with accuracy, (ii) early prediction for disease, (iii) patient oriented data with accuracy, (iv) The medical data, is securely stored and used in many places, (v) incomplete regional data are reduced and give the accuracy result. Goal of the concept is to choose the region and collects the hospital data or medical data of particular selected region, this process is using the machine learning algorithm. This term based on the data mining technique is used for disease prediction with accuracy. Then, finding the missing data based on latent factor to access the incomplete data and it is reduced. The previous system use the CNNUDRP (Unimodal Disease Risk Prediction), then continuously implements the next level use the CNN-MDRP (Mulimodal Disease Risk Prediction). The CNN-MDRP is overcome the drawback of CNN-UDRP.

The CNN-MDRP consists of the hospital data, that is structured and unstructured data. The CNNMDRP algorithm based prediction is more accurate, this accuracy is compared with previous system. The advantages of the concept is, better feature description and better accuracy, and the disadvantages of this system is, this feature is only applicable for the structured data so it is not good in disease description. Authors, In [2] have proposed in to the concept is machine learning based disease prediction using the big data for overcome the machine learning drawbacks. The smooth progress of big data is moves in the biomedical and healthcare communities in hospital for accurate results in any experiment result. This concept is (a) reduces the incomplete data and (b) effective disease prediction.

In [3] author has presented the data mining concept "Disease Prediction by using Machine Learning". The data mining best growth of the stage is develops that technique into the healthcare basis, the data analysis is an important part of every field. The data mining is predicts the information for healthcare is called rapid growth of medical care field. The existing one is designed the purpose of (i) analyze, (ii) manage, (iii) predict of healthcare data, it is described the overall healthcare systems. The concept of machine learning is applied into the disease-related information retrievals and the treatment processes in these types of process are achieved by using the data analysis. The predictions of outbreaks in diseases are using the decision tree, because it is very effective. This concept based experimental shows that result is related to the disease symptoms, so that data is described medical data using modified prediction model. If the concept choose the raining set like medical patient symptoms, than, use the decision tree, then, predicted, finally give the symptoms of patient and get the accurate result for disease prediction. This concept is only performs, that is predicts only the patient related information with low time and low cost. Authors, presents In [4] for "prediction of disease using machine learning over big data". Can develop the medical specialty basis this concept is applied to produce the medical data in to mass medical data, which means the data which is enlarged. The goal of this concept is targeted the simplest data is stored into the space of medical massive data analysis, called "medical data analysis in massive collection". It produces the accuracy and it reaches the 4.8% speed faster the CNN-UDRP. It only focuses this three data, (a) structured data, (b) text data, (c) structured and text data. In this proposed system is improves the medical data oriented term.

Concept presented by author, In [5] delivered theme is, "personalized disease prediction care from harm using big data", for healthcare analysis. This concept describes the medical field is a rich data industry because it holds the healthcare records, also. The daily treatment records are increased in every day that is it includes number of transactions, and the patient information is stored and retrieved from the database. The medical treatment records are every day updated one, because every day improves the patient health improvements based on treatment. It gives the correct solutions for different types of diseases. This system is change medical record, which means manually noted every medical oriented record into the electronic record that is, digitalize the medical care. This technology is simply

called, "e-healthcare". The medical data is stored in the database. The big data methods and the logics are used to analyze the statistical analytics. The proposed system is known as, "disease recommendation system", and this system holds the specialized tool, this tool is creating the profile. The profile making needs some information from the personalized persons, that is doctors, patients, etc, If entering the required field of the system finally get the personalized model health profile, but this personalization includes huge number of profiling information and other data. Enthusiastically presents the author, give the information are collected by In [6] namely, "Use the Weighted Ensemble to Neural Network based Multimodal Disease Risk Prediction (WENNMDRP) and feature selection of Ant colony improved classifier for disease prediction over the big data concepts". This concept feature selection performance is gives the dataset, this data set making is one the significant task. The feature selection task is splits into level by level, (i) first, structured the normal and more explainable models, (ii) apply the concept knowledge and learning its performance, (iii) finally, ready to preparing the clean, that is clear the data. Then, the proposed concept is analyzes the feature selection difficulties for big data based data analytics, so resolving this complexity by using the Improved Ant Colony Optimization (IACO) technique. This technique is early solves the missing data problem in incomplete data, which means it before set the latent factor mode, also. But is not easily selects the best feature from the medical data.

Table 3 Comparison table

Paper Number	Technique	Advantages	Disadvantages
1	Multimodal Disease Risk Prediction (CNN – MDRP)	(i) Medical data analysis with accuracy, (ii) early prediction for disease, (iii) patient oriented data with accuracy, (iv) The medical data, is securely stored and used in many places, (v) incomplete regional data are reduced and give the accuracy result.	This feature is only applicable for the structured data so it is not good in disease description.
2	Map Reduce (MR) algorithm	(a) Reduces the incomplete data and (b) effective disease prediction.	It reaches the 94.8% with the normal speed but it is quicker than CNN-UDRP and then, it report give the information is disease occurrences possibilities.
3	Decision Tree	(i) analyze, (ii) manage, (iii) predict of healthcare data	It predicts only the patient related information
4	Big data	The goal of this concept is targeted the simplest data is stored into the space of medical massive data	It only focuses this three data, (a) structured data, (b) text data, (c) structured and text data.
5	Disease recommendation system	This personalized profile is based on current treatment and any other treatment is takes, if can use the same profile, cannot require again make the profiling.	It Takes more time
6	Use the Weighted Ensemble to Neural Network based Multimodal Disease Risk Prediction (WENN-MDRP)	(i) First, structured the normal and more explainable models, (ii) apply the concept knowledge and learning its performance, (iii) finally, ready to preparing the clean, that is clear the data.	It selects the best feasible, but not previously checks the possibility

3. CONCLUSION

In the survey discuss Disease predicts the hospital data by using the different data mining technique. This analyze 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 structured data and unstructured data. The concept fulfill 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. This paper gives the merits and demerits of the recent techniques and its capabilities are studied. This paper concludes that there is no effective method discovers for Achieving Disease Prediction. So, further approaches should overcome all the above issues. Further implementation has to be done in order to Achieving High Disease Prediction using machine learning algorithm.

REFERENCES

[1]. Shraddha Subhash Shirsath, Prof. Shubhangi Patil Disease Prediction Using Machine Learn.Over Big Data". I international Journal of Innovative Research in Science, Engineering and Technology, [2018]. ISSN (Online) : 2319-8753, ISSN (Print) : 2347-6710.

[2]. Vinitha S, Sweetlin S, Vinusha H, Sajini S. "Disease Prediction Using Machine Learning Over Big Data". Computer Science & Engineering: An International Journal (CSEIJ), Vol.8, No.1, [2018].DOI: 10.5121/cseij.2018.8101.

[3]. Sayali Ambekar and Dr.Rashmi Phalnikar. "Disease Prediction by using Machine Learning". International journal of computer engineering and applications, Volume XII, special issue, May 18. ISSN: 2321-3469.

[4]. Lohith S Y, Dr. Mohamed Rafi. "Prediction of Disease Using Learning over Big Data - Survey". International Journal on Future Revolution in Computer Science & Communication Engineering. ISSN: 2454-4248.

[5]. J. Senthil Kumar, S. Appavu. "The Personalized Disease Prediction Care from Harm using Big Data Analytics in Healthcare". Indian Journal of Science and Technology, vol 9(8), DOI: 10.17485/ijst/2016/v9i8/87846, [2016]. ISSN (Print): 0974-6846, ISSN (Online): 0974-5645.

[6] Gakwaya Nkundimana Joel, S. Manju Priya. "Improved Ant Colony on Feature Selection and Weighted Ensemble to Neural Network Based Multimodal Disease Risk Prediction (WENNMDRP) Classifier for Disease Prediction Over Big Data". International Journal of Engineering & Technology, 7(3.27) (2018) 56-61.

[7] Asadi Srinivasulu, S.Amrutha Valli, P.Hussainkhan, and P.Anitha. "A Survey on Disease Prediction in big data healthcare using extended convolutional neural network". National conference on Emerging Trends in information, management and Engineering Sciences, [2018].

[8] Stephen J.Mooney and Vikas Pejaver. "Big data in public health: Terminology, Machine Learning, and Privacy", Annual Review of public Health [2018].
[9] Smriti Mukesh Singh, Dr. Dinesh B. Hanchate. "Improving Disease Prediction by Machine Learning". eISSN: 2395-0056, p-ISSN:2395-0072. [10]. Joseph, Nisha, and B. Senthil Kumar. "Top-K Competitor Trust Mining and Customer Behavior Investigation Using Data Mining Technique."Journal of Network Communications and Emerging Technologies (JNCET) www. jncet. org 8.2 (2018).
[11]. Kumar, B. Senthil. "Adaptive Personalized Clinical Decision Support System Using Effective Data Mining Algorithms." Journal of Network Communications and Emerging Technologies (JNCET) www. jncet. org 8.1 (2018).
[12]. Unnikrishnan, Asha, and B. Senthil Kumar. "Biosearch: A Domain Specific Energy Efficient Query Processing and Search Optimization in Healthcare Search Engine." Journal of Network Communications and Emerging Technologies (JNCET) www. jncet. org 8.1 (2017).
[13]. Kumar, B. Senthil. "Adaptive Personalized Clinical Decision Support System Using Effective Data Mining Algorithms." Journal of Network Communications and Emerging Technologies (JNCET) www. jncet. org 8.1 (2017).
[14]. Kumar, B. Senthil. "Data Mining Methods and Techniques for Clinical Decision Support Systems." Journal of Network Communications and Emerging Technologies (JNCET) www. jncet. org 7.8 (2017).