

## A Survey on Disease Prediction by Machine Learning over Big Data from Healthcare Communities

K.DeepthiKrishnan<sup>1</sup>, B.Senthil Kumar<sup>2</sup>

*M.Phil Scholar, Department of Computer Science, Sree Narayana Guru College, K.G.Chavadi, Coimbatore, Tamil Nadu, India<sup>1</sup>*

*Associate Professor, Department of Computer Science, Sree Narayana Guru College, Coimbatore, Tamil Nadu, India<sup>2</sup>*

*Corresponding Author: K.DeepthiKrishnan*

**Abstract:** The Enlarged and enhanced concept of big data is extracted in the medical field and the new concept is introduced in the paper concept. The survey concept is take the machine learning based disease prediction from medical field and uses the big data concept, which means the machine learning is a data mining techniques but this technique applied in disease prediction to come some difficulty such as, incomplete data, not suitable in large or big hospital and the some results are inaccurate, so this some difficulty are come in the existing, then, it will be move into the next level called the “Big Data”. The big data is large and huge data sets that holds, so this difficulty is overcome. The paper concept is machine learning based disease prediction over the big data. The big data is directly collects the information in Healthcare communities, because the big data is like, very knowledgeable concept. The proposed system aim is to (i) Analysis the optimal and accurate results on medical data, (iii) fast disease prediction (early predicts) from analyzed the data for medical field (hospital medical data), (ii) take the incomplete data and it move into the next level called “complete data” for disease analysis. The concept introduce the techniques namely, Multimodal Disease Risk Prediction (CNN-MDRP) based on Convolutional Neural Network. The paper additionally, describes the term Unimodal Disease Risk Prediction (UDRP) and it compares and analyze with the performance.

**Index Terms-** Big data analytics, Machine Learning, Healthcare, Disease Detection, Medical Data Analysis.

Date of Submission: 16-10-2018

Date of acceptance: 31-10-2018

### I. INTRODUCTION

The new trend and currently first preference concept of big data, it is used in many fields, which means it is extremely large data sets that it holds and it is analyzed the direct patterns, logical trends and associations, it specially related to human oriented behavior and interactions also accepted. So, this special and amazing concept is applied in to medical data and medical data oriented analyzing fields like that this paper concept. The development of big data is based on this concept and it is technique is applied in to biomedical field, healthcare communities, and medical data analysis with accuracy. The big data is used in this medical field get the advantages like, (i) early disease detection, (ii) patient care, (iii) community oriented services. In the time of medical data incompleteness, to uses the big data technique by reducing the analysis accuracy. If use the different regions that express the unique characteristics of the particular regional diseases, that make the weak prediction, so outbreaks the disease. In this proposed concept is basically based on the existing system focusing concept but it moved in to this proposed concept because existing systems disadvantages is overcome this current concept.

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 chronic disease outbreak. This chronic disease outbreak is solved in disease-frequent communities. But in this prediction model is use to occur the difficult in incomplete data. So, the incomplete data use the latent factor model so automatically reconstructing the prediction data finally it gets the missing data. But this type of chronic disease prediction is show the experimental result is infarction, which means blocking the blood supply area.

The proposed concept is used the new Convolutional Neural Network (CNN) concept, which means this concept include the two terms, and this term based concept is applied. This concept is connected with the brain, particularly cerebrum it mainly connected this concept especially, gyrus and sulcus with comparison. The main proposed concept is used the CNN based Multimodal Disease Risk Prediction (CNN-MDRP) algorithm. The CNN-MDRP is used in to the medical field, like that hospital data. The hospital data means (i) structured data, (ii) Unstructured data. The big data is give the best knowledge of out proposed concept, but the existing system proposes both types of data prediction in medical area and that basic development of medical big data

analytics. This proposed concept performance is calculated by comparing the different types of predicting algorithmic techniques. The proposed system is moreover reaches the 94.8% of high speed. Finally, the CNN-MDRP is faster than CNN-UDRP (Unimodal Disease Risk Prediction).

**KEY TERMS:**

**HEART DISEASE:**

The heart is an efficient part of our body. Life's, every moment is based on heart of our body, so the heart is not works correctly affects all parts of the body. The heart problem mainly affects the human parts such as brain, kidney, etc. It is nothing but pump, that is blood transferring area; the blood circulation is affect overall body parts are mainly affected. The blood circulation process is force to any problem, mostly suffers the organs that is brain it is related and connected the heart, so automatically suffers the heart.

**Symptoms of a Heart Attack:**

Symptoms of heart attacks are list out the following terms, (i) Pressure, (ii) Discomfort, (iii) Pain in the chest, (iv) arm. Some common indications are listed: (a) sweating, (b) nausea, (c) vomiting, (d) dizziness, then, it includes the anxiety, shortness of breath, extreme weakness, irregular heartbeats or rapid.

**Types of Heart diseases:**

Heart disease is based on all disease, the diseases are come in various stressing components of the heart, the heart denotes "cardio", which means entire heart disease is based on cardiovascular diseases.

**1. Coronary heart disease:**

Coronary Heart Disease (CAD) is one of the heart diseases, it is one of the most common diseases, it affects the blood circulations. This disease based on blood transfer oriented diseases.

**2. Arrhythmias:**

It is related with the disorder and recurring movement of the beating heart, that is heart beats levels are changed. The heart beat levels are slow, fast, irregular. The heart usually beats not created the problem, but the usually beats the heart its result is makes the short or small circuits in heart electrical system.

**3. Congestive heart failure:**

It is constraint or rules, nothing but cannot pump enough the blood to rest of body. This problem is occurs in the heart that is known as, "Heart Failure".

**4. Congenital heart disease :**

This disease is basic reason is heart defect. It refers the abnormal condition of heart. The heart functions are affected. This defect totally affects the heart works. This disease mostly affects the children.

**5. Cardiomyopathy:**

This type of disease is coming reason is Weak heart. That is the heart muscle and the changing the structure of the muscle, which affect the heart pumps.

**6. Angina pectoris:**

This types of disease is affects the heart, but actually it is chest pain. The blood transfers of the heart is failure or corrupted this disease is affected this term is called angina; another namely it referred to heart attack. The chest pain interval is few seconds or minutes.

**7. Myocarditis:**

It irritates the heart muscles, by the reason is (i) viral, (ii) fungal, (iii) bacterial contagion. This disease is rare disease but easily spreads, it is uncommon disease, which indicates the, (a) joins pain, (b) leg swelling, (c) fever, but it cannot directly affects the heart.

## **II. LITERATURE REVIEW**

In this paper [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 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 get the incomplete data and it is reduced. The previous system use the CNN-UDRP (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 is uses the hospital data, that is structured and unstructured data. The CNN-MDRP algorithm based prediction is produce 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 this paper [2] has 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.

The proposed concept is tested or experimented the real-life hospital data collections such as hospital oriented information like daily updated data- doctor data prefer doctor details, patient data prefer patient details, disease data prefer disease oriented data, etc.. This technique overcomes main two difficulties in the existing system are, (i) incomplete data, (ii) missing data. So rebuild the latent factor model. The concept is get the information from hospital that collected information forum called “structured and unstructured data”, and by using the Machine Learning Decision Tree algorithm and Map Reduce (MR) algorithm. The MR algorithm is used for data partitioning. 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.

The paper [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 the survey paper [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, this paper [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.

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 increased the computational time, so frame the time for clinical purposes. This concept is extracted and applies the application like, Collaborative Assessment and Recommendation Engine (CARE). The CARE is analyzes the performance limitation, and it improves the personalized disease prediction. The concept improved application CARE is classified into two types namely, (A) Current CARE Architecture, (B) Parallel CARE Architecture. The CARE is expresses the overall performance for patient oriented big data. It takes more time.

Enthusiastically presents the author, give the information are collected by the paper [6] namely, “Use the Weighted Ensemble to Neural Network based Multimodal Disease Risk Prediction (WENN-MDRP) 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.

The second technique WENN-MDRP is called the unheard technique, in this technique is helps to select the best features from medical data. These two methods are combined and give the special merit is improved prediction with accuracy, if this accuracy is evaluates to compare the experimental techniques. This concept is works only the time full fill the needed instances like, (i) accuracy, (ii) precision, (iii) recall. It selects the best feasible, but not previously checks the possibility.

Paper [7] gives the survey for Disease prediction in big data healthcare using extended CNN. This concept is applied in the medical field to implements the hospital. It provides the (i) high accuracy, (ii) high performance, (iii) high convergence speed. To select the particular region and then, analyzed the chronic diseases, that holds the structured data (extracted useful features), the unstructured data is use the CNN technique, so automatically selects the features. The novel CNN is proposed the medical data, and disease risk model is combined this data. The characteristic behavior of this system is selects the data via previous term. This term is previously applied is possible but not satisfied the disease changes, because disease level is not standard, it is changed in every seconds. To take the selected data from large number of data and improves the accuracy by using risk classification term. The proposed system aim is to predict the risk in liver oriented disease. So, the hospital dataset is related to the liver oriented disease and it collects only the structured data from liver disease information. In the proposed system is use the disease risk modeling and get the accuracy. But the risk prediction is depends on the different feature of medical data with higher accuracy.

This paper [8] author has presented the big data techniques in public health like, “Terminology, Machine Learning, Privacy”. The digitalize world is day by day increases the huge amount of data and increasing the data rate, so meet the staggering, but this problem is solved by incoming the new and fantastic forum and clever concept called the “Big Data”. The big data cannot lock the novel approach chances it understands the public health. The concept is express step by step but it is very force, firstly, it takes the classification of sources of the data like, big data, and then to clarify the terminology and then, identify its threads. The medical field oriented researches takes the big data, and it including the protection, hypothesis related generating researches takes the big data information. The interpretability is not aim this proposed system using machine learning technique.

This paper [9] author has presented the concept is, “Improving disease prediction by machine learning”, that is using the machine learning and improving the disease prediction. The big data is expanding the medical data, so improving this type of information. This concept use the genetic algorithm, it is utilizing the recover data, that is the missing data, then, it dataset includes the medical data. In this system using the two calculation terms namely, (i) KNN, (ii) SVM. The chronic diseases every increasing the data CNN-MDRP technique use the medical data. The database includes the medical data, and personal data and detailed history of patient is stored. The RNN based techniques are easily find out the logical data. This system uses the online and offline methods.

In the paper [10] author has presented the concept is, “Competitor Mining and Unstructured Dataset Handling Technique”, which used in healthcare communities. This paper competitive mining is describes with its related works. Finally gave competitor mining algorithms with its advantages and drawbacks. This paper experimental result shows CMiner++ yielded least computation time when comparing others.

Paper [11] gives the survey for Personalized Clinical Decision Support System Using Effective Data Mining Algorithms. This paper effectively designed a framework which is called A-CDSS (adaptive clinical decision support system). This framework manage and solved many research issues such as feature selection, classification issues like class Inequality and accuracy problems and finally the decision selection based on the diagnosed result. This paper result shows achieve the higher accuracy and effective clinical decision support system.

Author Senthil Kumar proposed [12] BioSearch engine with effective data mining machine learning algorithm with less energy for query processing. This proposed approach contains predictive data caching techniques which gives for fast and effective data retrieval. This system also integrates with auto query incremental algorithm to ease the search. Finally retrieved data's are ranked and summarized using RII result shows to the user view.

Paper [13] analyzes Clinical Decision Support Systems methods and techniques. This paper provides a review of different techniques and methods such as Genetic Algorithms, fuzzy, k-nearest neighbor, back propagation algorithm and ANN for clinical decision support system with its merits and demerits. This paper survey exposed that the Artificial Neural Networks (ANN) and fuzzy classification rules using data mining

techniques can incorporate data from many clinical and laboratory variables to provide better diagnostic accuracy in various clinical dataset.

The paper [14] author has presented the data mining concept for Clinical Datasets Using Weighted Genetic PCA Methods. Diagnosis and detection of diseases from patient electronic health records are very dynamic in nature and achieving that is a very promising area of research. So this paper implements a new weighted Genetic based algorithm with the use of effective weighted features from the PCA. The system finds the type2 Diabetes and Heart disease Classification using WGA technique. The system developed with the intension of high accuracy and less training overhead.

The paper [15] author has presented Identification of Diabetes Risk Using Machine Learning Approaches. With the numerous sizes in digital Healthcare data processes, the classification and prediction based on the statistical data is very tough .This survey discusses several machine learning approaches such as supervised learning, clustering and regression for Diabetes Risk this paper shows the advantages and disadvantages of several traditional classification algorithms based on different techniques .

In the paper [16] author has presented the concept is Feature Selection Methods Feature Selection Methods. The problem of feature selection, or choosing the most relevant features out of what can be an incredibly large set of data, is particularly important for accurate text categorization. So that this paper extract a feature vector for each new document by using feature weighting and feature selection algorithms for enhancing the text classification accuracy. After that we train our classifier by Naïve Bayesian (NB) and support vector machine (KNN) algorithms. In Experiments, although both algorithms are show acceptable good results for text classification.

**Table 1.0.** Comparison table

<b>Paper Number</b>	<b>Technique</b>	<b>Advantages</b>	<b>Disadvantages</b>
1.	Multimodel 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.
7.	Using extended CNN	(i) high accuracy, (ii) high performance, (iii) high convergence speed	The risk prediction is depends on the different feature of medical data

8.	Big data techniques	To clarify the terminology and then, identify its threads.	The interpretability is not aim this proposed system using machine learning technique.
9.	(i) KNN, (ii) SVM, Genetic algorithm.	This concept use the genetic algorithm, it is utilizing the recover data, that is the missing data, then, it dataset includes the medical data	Take more time.
10.	CMiner++	To predict the medical data.	To predict the unstructured data is very complex one.
11	A-CDSS Model	This achieves effective clinical decision support system.	Discovered features more difficult process.
12	BioSearch	Less energy for query processing.	Need Effective Ranking algorithm to view result
13	Artificial Neural Networks (ANN)	Provide better diagnostic accuracy.	Effective feature selection is complex task.
14	WGA	high accuracy and less training overhead	Need more training sample.
15	clustering and regression model	This achieves effective prediction.	Prediction based on the statistical data is very tough.
16	KNN	Achieve Accurate relevant features.	Computation cost is quite high.

The above table 2.0. Depicts the working methodologies of various big data techniques which can be used to achieve machine learning based prediction of disease.

### III. 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 two types namely, (i) structured data, (ii) 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 (WENN-MDRP) 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". e-ISSN: 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).
- [15]. Sreejith, B. Senthil. "Identification of Diabetes Risk Using Machine Learning Approaches." *Journal of Network Communications and Emerging Technologies (JNCET)* www. jncet. org 7.8 (2017).
- [16]. Bhavitha Varma, B. Senthil. " A Different Type of Feature Selection Methods for Text Categorization on Imbalanced Data." *Journal of Network Communications and Emerging Technologies (JNCET)* www. jncet. org 8.1 (2017).

K.DeepthiKrishnan. " A Survey on Disease Prediction by Machine Learning over Big Data from Healthcare Communities." *IOSR Journal of Engineering (IOSRJEN)*, vol. 08, no. 10, 2018, pp. 53-59.