7th National Conference on
"Role of Engineers in Nation Building" organized by VIVA Institute of Technology, Mumbai
(1st and 2nd March 2019)
International Journal of Engineering Research & Science (IJOER)
ISBN: [978-93-5321-411-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2019]

Survey on E-Doctor: An Innovative approach for Disease Prediction

Chinmay Devdhekar[1], Kaushik Patil[2], Krutika Diwale[3], Reshma Chaudhari[4]
Department of Computer Engineering, Mumbai University, INDIA

Abstract— *In the recent years there has been tremendous growth in Data Mining and Machine learning in the fields of Biological Data Analysis, Financial Data Analysis, Telecommunication industry and many more other industries and businesses. The main objective of this paper is to get detailed information on Clinical and health care applications and perform a survey on the applications that are proposed in the paper which uses data mining and machine learning technique to predict accurate result. This paper also gives us an insight about Clinical and Biological data that provides a detailed discussion of data mining techniques that can improve various aspects of Clinical data Prediction. The machine learning algorithms are used to extract new patterns from large dataset and knowledge associated with this patterns are also extracted. It also gives a detail study on various technique uses in data mining and evaluating the best technique for classifying or predicting disease and shows that ANN is best among all data mining technique.*

Keywords—Biological data analysis, Data mining, E-Doctor, Prediction, Machine Learning.

I. INTRODUCTION

Data Mining is that the mining of information from a large data, or discovery of valid data in terms of patterns or rules from huge amounts of information. Data mining is closely associated with Knowledge Discovery in Databases (KDD). The techniques of KDD includes: pattern recognition, clustering, classification tree, case-based reasoning, AI techniques, statistics, neural networks and others

Extraction of knowledge isn't the sole method we'd like to perform; data mining additionally involves alternative processes like information cleansing, data Integration, data transformation, data processing, and Pattern analysis and data presentation. Once of these processes are over, we might be ready to use this data in several applications like Fraud detection, Marketing, Research, Production Management, Science Exploration, etc.

One of the most important issues in data mining in medication is that the raw medication information is voluminous and heterogeneous. This knowledge may be gathered from numerous sources like from oral communication with patients, science laboratory results and interpretations of doctors. E-doctor prediction systems are designed to help the patient and also the doctors for Check-up and diagnosing. This technique provides immediate facilitate from the doctor on their health problems by victimization online health care system. It assists patient to predict the disease he or she should be affected by, and helps doctors to keep the record of the patient.

II. LITERATURE SURVEY

2.1 Prediction of heart Disease using hybrid technique in data mining classification [1].

In this paper a standout amongst the best order procedure is utilized that is Back Propagation strategy which is utilized for non-direct connections. This method has the disadvantage of being stuck in nearby minima causing not ready to accomplish most extreme benefit of the system. To beat this issue utilization of a standout amongst the best enhancer strategy for example Genetic algorithm is utilized which utilizes the wonders of transformation and hybrid over different ages accomplishing 75% of precision. An end can be made that neural system is best among all the arrangement systems when we talk about expectation. Or on the other hand grouping of a non-straight information. Back Propagation calculation which is the best classifier of Artificial Neural Network utilizes the refreshing method of weights by correcting the error in reverse can be utilized or on the other hand grouping of unrelated information. BP calculation which is the best classifier of Artificial Neural Network which utilizes the refreshing method of loads by engendering the error in reverse can be utilized

International Journal of Engineering Research & Science (IJOER)
ISBN: [978-93-5321-411-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2019]

2.2 Prediction in heart Disease using technique of data mining [2].

In this paper the framework utilizes Naïve Bayer Technique for taking care of substantial measure of information however .Naïve Bayer procedure is inclined to have Loss of precision yet ANN (Artificial Neural Network) makes a valuable device to assist specialists with analyzing, demonstrate complex clinical information. Neural Networks can without much of a stretch handle missing or commotion information. When prepared, does not have to reconstruct. It can without much of a stretch work with substantial number of datasets.

2.3 Knowledge discovery analysis and prediction in Healthcare using Data Mining and Analytic [3].

In this paper the creator has utilized the neural system to play out a vital job in the field of prescription. It very well may be utilized to analyze different infections which could be conceivable by giving distinctive impacts or changes which are felt in body by allocating weights to every parameter. It can likewise be utilized to locate the suitable medication containing the required medication for a specific disease. Likewise it tends to be utilized to separate between the patients who required month to month registration and alternate patients. It can likewise be utilized for image processing. Neural Network framed thusly is prepared for the unknown information and it will respond based on recently gained learning .neural systems are perfect for multiprocessor frameworks, where countless activities are performed in parallel

2.4 A review on using various Data Mining techniques for Heart disease [4].

This proposed system contains correlations and results of various evaluation techniques. Thus helping the medical professional to take up the correct choice on opportune time giving suitable treatment. In this framework, with the help of Genetic algorithm the weights utilized in Back Propagation are improved at first and it is feed into system as info. An enhancement execution is acquired by utilizing Genetic algorithm. In some cases, 100% exactness with Neural Network. Data mining system help in looking through some clear information from this gigantic record, in this manner sparing the time and Predictions can turn out to be progressively precise in helping specialists to settle on choice for the most part in analysis.

2.5 Disease forecasting system Using Data Mining Methods [5].

This proposed framework contains comparisons & results of various evaluation strategies. Thus helping the medical professional to take up the correct choice on perfect time giving suitable treatment. In this framework, with the assistance of Genetic calculation the loads utilized in Back Propagation are advanced at first and it is feed into system as information. An enhancement execution is gotten by utilizing Genetic calculation. In some cases, 100% precision with Neural Network. Data mining method help in looking through some clear information from this huge record, consequently saving the time and Predictions can turn out to be increasingly exact in helping specialists to settle on choice for the most part in determination

2.6 Seminal Quality Prediction using Optimized ANN and genetic algorithms [6].

In this paper the creator has proposed ANN algorithm which is an powerful data mining instrument that can be utilized for accomplishing the objective and this paper uses genetic algorithm to advance the structure of ANN to classify samples. The quantity of neurons in the main layer is equivalent to dimensional of highlight vectors. The quantity of neurons in keep going layer depends on number of classes in class issue. The algorithm begins with some random chromosomes. Every chromosome represents architecture for neural system and related learning rate. The total dataset is partitioned in two subset preparing and test subset. The investigations are done on red fertility analysis dataset. The technique beats SVM a choice tree and Naïve Bayer.

2.7 Analysis & Prediction of Breast Cancer & diabetes disease dataset using Data Mining Technique [7].

In this paper, Two disease dataset gathered from uci machine learning repository. First dataset is breast cancer dataset having 10 attributes and second dataset is diabetes dataset having 9 attributes. Naïve Bayer, SMO, REP Tree, J48 and MLP calculations are utilized to classify breast cancer and diabetes dataset on WEKA interface. The exhibitions of these five calculations have been broke down on breast cancer and diabetes dataset utilizing preparing information testing mode. J48 gives 74.28%accurate outcomes as comparison to other calculation on breast cancer dataset and SMO gives 76.80% precise outcomes on diabetes dataset.

International Journal of Engineering Research & Science (IJOER)
ISBN: [978-93-5321-411-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2019]

2.8 An Empirical study on prediction on heart disease using classification data mining techniques [8].

In this paper the author has clarified the limitations of regular medicinal frameworks handled in the order demonstrate which recognize complex non-linear relationship between factors. Algorithm with most prominent precision is basic for order of coronary illness design for the clinicians to analysis it. It is seen that Naive Bayer has a more prominent precision when contrasted with K – Nearest neighbors', Decision-Tree and Neural Network. Enhancement in these frameworks would help the quality of medical diagnostics decisions..Data mining is sorting the information for its powerful and effective use as according to its critical value and need to be accessed. The multi-parametric component with linear and nonlinear variable characteristic of heart rate variability precisely for the supine, left lateral and right lateral position. The process of knowledge acquisition and need to collect adequate data was the major drawback in the traditional analysis to mining the data, sorting variables and their effect.

2.9 Study of Machine Learning Algorithms for Disease Prediction using Principal of Component Analysis [9].

In this paper It combines statistical analysis, ML and database technology to extract hidden patterns and connections from databases. DM Algorithm is utilized for confirming the precision in predicting diabetic status .Classification is one of the most repeatedly studied problems by DM and ML specialists. Naive Bayer and Decision tree has been applied with and without utilizing PCA on the dataset. It utilizes PCA to diminish the quantity of attributes. After reducing the size of the dataset, SVM beats Naïve Bayer and Decision tree

2.10 Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction [10].

In this paper, the numbers of experiments are conducted to compare the performance of predictive data mining technique on the same dataset and the outcome revels that Decision tree and Bayesian Classification having same accuracy but other predictive methods like KNN, Neural Networks, Classification based on clustering is not performing well. Decision Tree is a popular classifier which is simple and easy to implement. It requires no domain knowledge or parameter setting and can handle high dimensional data. Naive Bayer is a statistical classifier which assumes no dependency between attributes. It attempts to maximize the posterior probability in determining the class. Accuracy of the KNN algorithm can be severely degraded by the presence of noisy or irrelevant features, or if the feature scales are not consistent with their importance.

2.11 Intelligent Machine for Nutritional Management and Healthcare using ANN [11].

In this paper the database with respect to supplements is gathered; a PC program has been developed dependent on neural system. Five parameters, day by day protein, vitality and water necessities, are determined through prepared artificial neural systems utilizing a database of people. Then the program was giving higher mistake values for large number of sources of input. The primary purpose for this was showed to end that on the off chance that number of sources of info is not exactly number of yields, at that point mistake will be high. Additionally hidden neurons must be expanded to get increasingly precise outcomes.

III. ANALYSIS TABLE

The below table is the summary of the studied research papers and the different techniques used on analysis of different disease prediction.

TABLE

Sr. No.	Title of Paper	Technique/ Methods	Accuracy
1.	Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification [1]	Neural Network & Genetic algorithms	75%
2.	Predictions in Heart Disease Using Techniques of Data Mining [2].	Neural network	-

International Journal of Engineering Research & Science (IJOER)
ISBN: [978-93-5321-411-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2019]

3.	Knowledge Discovery Analysis and Prediction in Healthcare using Data Mining and Analytics [3].	Neural Network	ANN performs better
4.	A review on using various Data Mining techniques for Heart disease [4].	ANN Algorithm	89%
5.	Disease forecasting system Using Data Mining Methods [5].	C4.5	79%
6.	Seminal Quality Prediction using Optimized ANN and genetic algorithms [6].	ANN Algorithm	93%
7.	Analysis & Prediction of Breast Cancer & diabetes disease dataset using Data Mining Technique[7].	Naïve Bayer SMO Algorithm	Naïve Bayer : 72% SMO: 76%
8.	An Empirical study on prediction on heart disease using classification data mining techniques[8].	Naïve Bayer Algorithm	83.7%
9.	Study of Machine Learning Algorithms for Special Disease Prediction using Principal of Component Analysis[9].	SVM Algorithm	SVM outperforms Naïve Bayer & Decision tree
10.	Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction [10].	Naïve Bayer, KNN, Decision tree	Naïve Bayer:52.33% Decision tree: 52% KNN: 43.67%
11.	Intelligent Machine for Nutritional Management & Healthcare using ANN [11].	ANN	ANN gives better result

IV. CONCLUSION

Through the detailed study of clinical data and the data mining technique used on this data we could come to the conclusion that the technique used such as K-means clustering algorithm, Naïve Bayes, SVM algorithm are giving less accurate results on large data set of medical data. Whereas Neural Network with the use of genetic algorithm gives more accurate result then the other techniques. Clinical data can be much more complex and predicting result on the bases of unrefined data can be difficult. Therefore diseases which are identified have less accuracy. With the more refined dataset technique Decision tree and Naïve Bayer can also give more accurate results. This problem is overcome by using a system that uses neural network algorithm which uses the technique of Back Propagation to adjust the weight. The system using neural network on unrefined data will give much better result.

REFERENCES

[1] A.Dewan, M.Sharma "Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification" 2015- 2nd International Conference on Computing for Sustainable Global Development.

[2] M.Gandhi, Dr.S.N.Singh, "Predictions in Heart Disease Using Techniques of Data Mining", Proceeding of the 2015-1st International Conference on Futuristic trend in Computational Analysis and Knowledge Management.

[3] A. Raul, A.Patil, P.Raheja, R.Sawan, "Knowledge Discovery, Analysis And Prediction in Healthcare using Data Mining And Analytic", 2016-2nd International Conference on Next Generation Computing Technologies (NGCT-2016) Dehradun, India 14-16 October 2016.

[4] D.C Bindushree "A review on various data mining technique for analysis of heart disease prediction"978-1-5386-0569-1$31.00 c 2017 IEEE.

[5] M.A.NisharaBanu, B.Gomathy,"Disease Forecasting System Using Data Mining Methods"2014 International Conference on Intelligent Computing Applications.

International Journal of Engineering Research & Science (IJOER)
ISBN: [978-93-5321-411-1] ISSN: [2395-6992] [Vol-5, Issue-3, March- 2019]

[6] A.A.Bidgoli, H.E.Komleh, S.Mousavirad, "Seminal Quality Prediction using Optimized Artificial Neural Network with Genetic Algorithm".

[7] D.Verma, Dr.N.Mishra. "Analysis and Prediction of Breast cancer and Diabetes disease datasets using Data mining classification Techniques." Proceedings of the International Conference on Intelligent Sustainable Systems (ICISS 2017).

[8] T.J. Peter "An Emperical study on prediction of Heart disease using classification of data mining technique'' IEEE International Conference on. 2016.

[9] M.Kishor, "Study of Machine Learning Algorithms for Special Disease Prediction using Principal of Component Analysis (2014)."2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication.

[10] Eman AbuKhousa, Piers Campbell "Predictive Data Mining to Support Clinical Decisions: An Overview of Heart Disease Prediction Systems" International Conference on Innovations in Information Technology (IIT).

[11] A.S.Pawar Dr.A.R.Agarkar "Intelligent Machine for Nutritional Management &Healthcare using ANN"2015th International Conference on Computing communication control and Automation.