

A Literature Review on Various Disease using Data Mining and Machine Learning Techniques

Sheetal Vijayvargiya¹, Priya Jyotiyana², Madhu Choudhary³

Assistant Professor, Department of Computer Science & Engineering^{1,2,3}

Jaipur Engineering College and Research Centre, Jaipur, Rajasthan, India

Abstract: A huge number of data mining tools have been developed over the decades, due to widespread and complexity involved in building the data mining applications. Many tools use various algorithm base and methods to carry out many data mining tasks. Every tool has its own pros and cons. Data mining gives an efficient way to retrieve the required data or information from raw and heterogeneous data. There are many methods to predict the medical disease with high accuracy and classification. There are many challenges and issues in data mining for applicable in medical practice. To extract the information and minimize the effort of an expert system for other clinical concern like drug addiction. This review paper aimed at evaluating and analysing a range of data mining tools and techniques for optimally predicting the numerous medical diseases to provide the healthcare section with high competence and more effectiveness. Data mining gives an efficient way to retrieve the required data or information from raw and heterogeneous data. There are many methods to predict the medical disease with high accuracy and classification. There are many challenges and issues in data mining for applicable in medical practice. To extract the information and minimize the effort of an expert system for other clinical concern like drug addiction.

Keywords: Data Mining, Machine Learning, Heart Disease, Feature Extraction, & Deep Learning

I. INTRODUCTION

Data mining will determine the most significant information from the data ware house of the organization, it is powerful technology, tool and knowledge. It is very essential step that collective study huge amount of habitually data. To search latest pattern in healthcare industry, there exist a variety of scalable and interactive data mining methods [1]. In present era, various data mining techniques are developed such as classification, pattern matching, data visualization; clustering and meta rule guided mining [2].

Data mining is broadly used in communication areas, credit review, stock market prediction, marketing, banking, education, health and medicine, forecasting, fraud detection, etc. but data mining hold important presence in each and every field of medical science for the analysis of various diseases such as heart disease, lung cancer, breast cancer, kidney stone liver disorder etc.

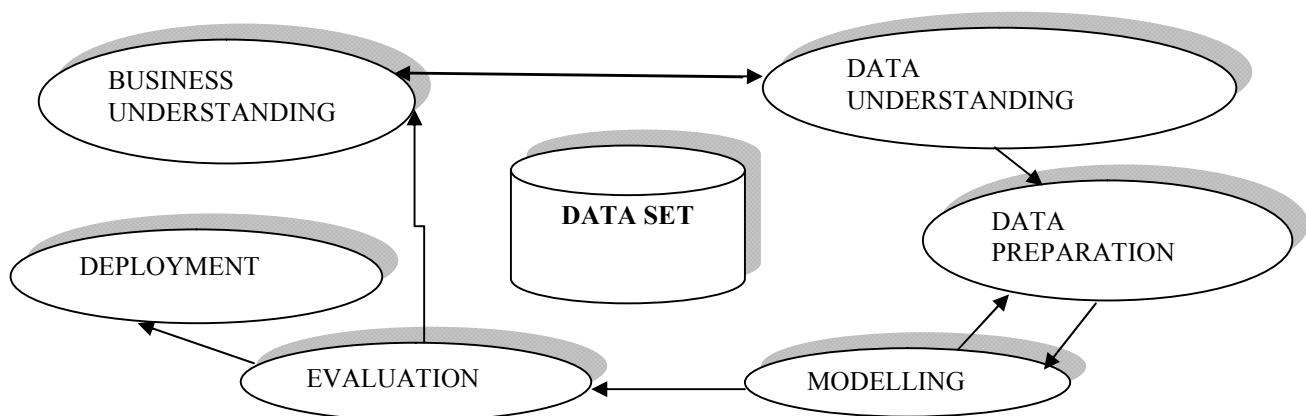


Figure 1: Data mining process model

II. BACKGROUND

With the unborn blow-up of information communication technologies, data mining will negotiate full eventuality in the knowledge discovery which is hidden in medical data and coincidentally retrieve the information and reduce the load of a complete system for various disease under consideration.

Within the matter of knowledge integrity evaluation, two biggest challenges are

1. For comparing the two data interpretation (ahead and later) develop the more effective algorithm.
2. For assessing the change in particular, develop the algorithm on statistical significance of individual patterns that are collected the data with the help of data mining algorithms.

III. RELATED WORK

Different functionalities proposed by different tools and ways are vaticination of data, pattern matching, categorization, characterization and conception of data. The development and significance of data mining algorithms necessitates the deployment of important software tools. As a result, a number of authors have recommended and used the diversity of data mining tools.

TABLE I: Literature review of diseases using data mining techniques

Author	Year of publication	Disease considered	Objectives	Accuracy	Techniques	Dataset
Uddin, Shahadat, et al.[4]	2019	Heart	Machine learning technique is applied for finding the important features and improving the accuracy for prediction of cardiovascular disease.	88.7%	Hybrid random forest with a linear model	UCI machine learning repository
Khourdifi Youness, and Mohamed Bahaj [5]	2019	Heart	Machine learning algorithms are used to improve heart disease prediction and classification	99.65%	K-Nearest Neighbour, Support Vector Machine, Naive Bayes, Random Forest	UCI machine learning repository
Kaur, Gaganjot, and Amit Chhabra	2014	Heart disease, Diabetes	Data mining algorithm are used for the prediction of diabetes contributes to heart disease and increases the effect of other organs includes kidney disease, nerve damage, blood vessel damage and blindness.	99.87%	J48 Decision Tree	Pima Indians Diabetes Data Set
Rady, El-Houssainy A., and Ayman S. Anwar	2019	stages of chronic kidney disease	Probabilistic neural network is used for better classification and prediction for determining the critical stages in chronic kidney disease.	96.7%	Probabilistic Neural Networks	61 chronic kidney disease patients
Singh, Navdeep,	2018	heart disease	To design a proposed model for heart illness	97.14%	GA_Fuzzy_Naive Baise	online dataset

Punjab Firozpur, and Sonika Jindal			recognition using data mining techniques that are fit for improving the accuracy of heart infection conclusion.			repository
Dahiwade Dhiraj, Gajanan Patle, and Ektaa Meshram	2019		k- nearest neighbor and convolutional neural network machine learning algorithm are used for prediction of general disease symptoms and accuracy.	84.5%	K-Nearest Neighbor (KNN) and Convolutional neural network (CNN)	UCI machine learning
Mohan, Senthilkumar, Chandrasegar Thirumalai, and Gautam Srivastava	2018	Heart	Prediction of cardiovascular disease hybrid random forest with linear model is used for finding the important features of disease.	88.7%	Hybrid random forest with a linear model	UCI Cleveland dataset
Kohli, Pahulpreet Singh, and Shriya Arora	2018	Heart, Breast	For the prediction of disease various classification algorithms are used and backward modeling using the p-value test are used for feature selection of each dataset.	87.1%, 98.57%	Logistic Regression, Adaptive Boosting	UCI repository
Ali, Farman, et al	2020	heart	Smart healthcare system is proposed for the prediction of heart disease using deep learning and feature fusion approaches.	98.5%	Ensemble deep learning model and feature fusion methods	Cleveland and Hungarian datasets.

IV. CONCLUSION

To reuse the raw data and give a new judgment towards the various conditions a machine literacy system is used. Heart complaint vaticination is gruelling approach in the medical field. On the other hand, the death rate can be significantly controlled if the complaint discovery can be done in early stage and apply the preventative and remedial measures as soon as possible. It's largely desirable to apply the analysis to the real-world dataset rather of theoretical approaches. In future this exploration can be done with various machine literacy ways to more vaticination of these styles and ways.

REFERENCES

- [1] Gil, A.G., Wagner, E.F., Vega, W.A. Acculturation, familism, and alcohol use among Latino adolescent males (Longitudinal relations) . J Community Psychol (john wiley & sons) Vol. 28, pp. 443–458, 2000.
- [2] S. H. Liao, P. H. Chu, and P. Y. Hsiao, "Data mining techniques and applications - A decade review from 2000 to 2011," Expert Syst. Appl., vol. 39, no. 12, pp. 11303–11311, 2012.

- [3] Bhatla, Nidhi, and Kiran Jyoti. "An analysis of heart disease prediction using different data mining techniques." *International Journal of Engineering* 1.8 (2012): 1-4.
- [4] Uddin, Shahadat, et al. "Comparing different supervised machine learning algorithms for disease prediction." *BMC medical informatics and decision making* 19.1 (2019): 1-16.
- [5] Khourdifi, Youness, and Mohamed Bahaj. "Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization." *International Journal of Intelligent Engineering and Systems* 12.1 (2019): 242-252.
- [6] Kaur, Gaganjot, and Amit Chhabra. "Improved J48 classification algorithm for the prediction of diabetes." *International journal of computer applications* 98.22 (2014).
- [7] Rady, El-Houssainy A., and Ayman S. Anwar. "Prediction of kidney disease stages using data mining algorithms." *Informatics in Medicine Unlocked* 15 (2019): 100178.
- [8] Singh, Navdeep, Punjab Firozpur, and Sonika Jindal. "Heart disease prediction system using hybrid technique of data mining algorithms." *International Journal of Advance Research, Ideas and Innovations in Technology* 4.2 (2018): 982-987.
- [9] Dahiawade, Dhiraj, Gajanan Patle, and Ektaa Meshram. "Designing disease prediction model using machine learning approach." 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2019.
- [10] Mohan, Senthilkumar, Chandrasegar Thirumalai, and Gautam Srivastava. "Effective heart disease prediction using hybrid machine learning techniques." *IEEE access* 7 (2019): 81542-81554.
- [11] Kohli, Pahulpreet Singh, and Shriya Arora. "Application of machine learning in disease prediction." 2018 4th International conference on computing communication and automation (ICCCA). IEEE, 2018.
- [12] Ali, Farman, et al. "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion." *Information Fusion* 63 (2020): 208-222.