International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 04 | Apr 2019 www.irjet.net p-ISSN: 2395-0072

A Literature Review on Heart and Alzheimer Disease Prediction

Jayashri J. Patil[1], Nilesh Vani[2]

[1]M. Tech-Student, Dept. of Computer Engineering, GF's Godavari of Engineering, Jalgaon, Maharashtra, India
[2]Assiatant Professor, Dept. of Computer Engineering, GF's Godavari of Engineering, Jalgaon, Maharashtra, India

---

Abstract - *Nowadays, heart diseases are very common and one of the major causes of death across the world. This calls for an accurate and timely diagnosis of the heart disease. There is abundant data available with the health care systems; however, the knowledge about the data is rather poor. Data scientists have attempted several methods in order to improvise the examination of large data sets. Previously, various data mining techniques have been implemented in the healthcare systems, however, the hybridization in addition to a single technique in the identification of heart disease shows promising outcomes, and can be useful in further investigating the treatment of the heart diseases. The framework enables the representation, extraction, and mining of high order latent structure and relationships within single and multiple disease sequences. This work attempts to survey some recent techniques applied towards knowledge discovery for heart disease prediction and further proposes a novel prediction method with improved accuracy.*

Unfortunately, Alzheimer's disease (AD) cannot be slowed or cured with today's medication. The studies have revealed that - a cognition drop is a precursor of AD, the progression of AD is highly correlated to cognition decline, and AD's early detection and intervention becomes increasingly clear to be the best choice of improving quality of life for persons with probable AD. This survey aims to improve the predictive model by focusing on AD early detection. Compared to models built from traditional approaches such as neuron networks, Bayesian networks, we propose a novel prediction method with improved accuracy.

Key Words: Heart Disease Prediction, Machine Learning, Data Mining, Alzheimer's disease (AD), deep learning, AD early detection.

1.INTRODUCTION

1.1 Data mining:

Data mining (sometimes knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

1.2 Basic terms related to data mining:

1.2.1 Classification

Classification is a data mining (machine learning) technique used to predict group membership for data instances. For example, you may wish to use classification to predict whether the weather on a particular day will be "sunny", "rainy" or "cloudy".

1.2.2 Supervised learning:

Supervised learning is the machine learning task of inferring a function from labeled training data. The training data consist of a set of training examples. In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal). A supervised learning algorithm analyzes the training data and produces an inferred function, which is called a classifier. The inferred function should predict the correct output value for any valid input object. This requires the learning algorithm to generalize from the training data to unseen situations in a "reasonable" way.

1.2.3 Unsupervised learning:

In machine learning, unsupervised learning refers to the problem of trying to find hidden structure in unlabeled data. Since the examples given to the learner are unlabeled, there is no error or reward signal to evaluate a potential solution. This distinguishes unsupervised learning from supervised learning.
The heart attack occurs when the arteries which supply oxygenated blood to heart does not function due to completely blocked or narrowed.

1.2.4 Prediction:

Models continuous-valued functions, that is predicts unknown or missing values.

1.3 Some Existing Heart Disease Prediction Algorithms :

The existing 11 algorithms based on the prediction of heart disease which has been used by various researchers are explained below:

Naïve Bayes (NB): A classification approach relied on Bayes hypothesis having a presumption of distinct factors [13]. The dataset is categorized in a distinct manner for predicting the heart disease occurrence. It is particularly used for very large data sets such as medical related data. To predict the heart disease through probability Naïve Bayes is used.

Decision Tree (DT): Decision tree is a type of supervised learning algorithm. It works for both categorical and continues input and output variables [13]. In the diagnosis of heart disease, decision tree will segregate the datasets based on all the values of the attributes and identify the attribute, which creates the best homogeneous sets of data. It also provides classified report for the heart disease.

K-Nearest Neighbor (KNN): For predictive problems such as classification and regression KNN is used [13]. KNN algorithm is mainly used to find the values of the factors of heart disease by using the k user defined value. By using the K values, it is possible to make boundaries for each class of heart disease related attributes.

Average K-Nearest Neighbor (AKNN): To make the KNN a faster algorithm AKNN is proposed by C. Kalaiselvi [12]. AKNN reduces the training sample size of heart disease dataset to n super samples. When the test samples of heart disease are given, the AKNN searches only the sample data and find the closest input thus making the KNN a faster algorithm.

Java Implementation of C4.5 algorithm (J48): J48(C4.5) is used to generate decision tree which is an extension of ID3 (Iterative Dichomester 3). For each heart disease dataset all the attributes are measured based on the type of chest pain. Based on the chest pain type the J48 algorithm grows an initial tree by using divide and conquers technique [7]. Reduces Error Pruning Tree (REP TREE): REP Tree is a speed decision tree algorithm. It constructs a regression tree using data gathering and removes it using reduced-error pruning.

Iterative Dichomester 3 (ID3): ID3 is the precursor to the C4.5 algorithms which is used to generate decision tree structure for the classified heart disease data. Different categories of the Heart disease data are investigated and every category are tested for heart disease prediction [5]. In each class the standard values are given to attributes such as Resting Blood Pressure, Serum Cholesterol, etc. In the tree structure each sub node represents the trained dataset of each class.

Support Vector Machine: SVM is a supervised machine learning algorithm which can be used for both classification or regression. It classifies the heart disease dataset by finding a hyperplane which separates the dataset into two classes. It performs better in terms of specificity and adaptability.

Artificial Neural Network (ANN): Artificial Neural Networks are the biologic outstanding act executed on the system to perform distinct functions such as classification, clustering, etc. The prediction of heart disease consists of the input layer, hidden layer, outputs and training classes which are based on the predicted class. It provides minimized error for the prediction of heart disease.

Fuzzy Analytical Hierarchical Process (Fuzzy AHP): Fuzzy AHP provides a hierarchical structure which shows the ultimate aim of the prediction of heart disease and its replacements are drafted. The local weight of each attribute and global weight of each attribute used in determining heart disease are calculated by using Fuzzy AHP. [2]

Feed Forward Neural Network (FFNN): The Feed Forward Neural Network is the first and simple type of artificial neural network except that the connections between the units do not form a cycle so that the information moves in a single direction. The given inputs are multiplied by weights, summed and combined by sigmoid activation to predict the output for the heart disease.

2. LITERATURE SURVEY

A survey is carried out on different data mining techniques and the different parameters used for prediction of heart disease are discussed. The accuracy obtained with these models are also mentioned.

2.1 For Heart Disease :

A. Hlaudi Daniel Masethe et al., (2014) used data mining algorithms like J48, NB and REP TREE for predicting heart attacks. The medical database was collected from the doctors in South Africa. The various attributes considered were Gender, age, CPT, ECG, RBP, Thalach, serum cholesterol, alcohol, obesity(diet) and smoking. WEKA - Waikato Environment for Knowledge Analysis tool was used for discovering, analysing and predicting patterns for heart disease. The accuracy obtained were 99.0741, 99.222, 98.148 for J48, REPTREE and NB respectively. [7]

Limitations: Some important attributes such as Col-Ves, Thal, Ex-Ang, etc., are not considered for prediction. Here various data mining algorithms are implemented and compared to find the best method for prediction. They came to a conclusion that algorithms such as J48 and REPTREE are efficient in the prediction of heart disease. This conclusion is derived by not considering some efficient Data mining algorithms such as Regression and Artificial Neural Network algorithms.

B. Theresa Princy R et al., (2016) predicted heart disease using ID3 and KNN algorithm. The ID3 algorithm is used as a classifier, KNN algorithm organizes and pre-processes the incorrect values which are considered as the training set. The basic factors along with some additional factors such as smoking were included. The accuracy level increased to 80.6%. [5]

Limitations: Very few factors are considered for the prediction. Some common influencing factors such as CPT (Chest Pain Type), RECG (Resting electrocardiographic (ECG)) and indirectly influencing factors such as Alcohol and Obesity are not considered. Without using these important factors, the prediction could not be given precisely.

C. VivekanandanTet al., (2017) proposed the challenging tasks of selecting critical features from the enormous set of available features and diagnosing heart disease. DE (Modified Differential Evolution) algorithm is used to perform feature selection. Prediction of heart disease was carried out using Fussy AHP and Feed-Forward neural network. Using 9 attributes an accuracy of 98% was achieved. [2]

Limitations: More number of inputs are not listed and Error minimization is not carried out properly. It could have been carried out by using effective back -propagation model. Without proper testing they have mentioned that large data sets can also be adopted.

D. Aditi Gavhane and Gouthami Kokkula (2018) proposed the system that use the neural network algorithm multi-layer perceptron (MLP) to train and test the dataset. Multi-layer perceptron algorithm is a supervised neural network algorithm in which there will be one layer for the input, second for the output and one or more for hidden layers between these two layers.

Limitations: The existing algorithm has been modified into new algorithm which takes only two values for prediction. By considering just two values they have come to a conclusion that their proposed algorithm achieves higher accuracy when compared to other algorithms.

2.2 For Alzheimer Disease:

A. Tingyan Wang andJason L. Qiu (2018) proposed the system that uses deep learning algorithm. In this, they explored how RNN models could be applied to AD early detection modeling (AD-EDM). In particular, using an LSTM RNN, they developed a predictive model based on the NACC's early stages' patient data.

Limitations: The dimensionality of the data is reduced using attribute selection method, since it consumes more time for classification. This decrease in the number of attributes does not give the correct prediction.

B . Veeramuthu et al. (2014) developed a CAD (AR mining algorithm) tool for decision making about the presences of abnormalities in human brain. The author suggested preprocessing of PET dataset for instance, spatial normalization and intensity normalization. Fisher Discriminants ratio (FDR) was used for feature extraction to get ROIs. The instances were classified to normal if the extracted number of verified rules were above the final threshold otherwise image was classified as AD.

Limitations: No dataset details, missing values or any pre-processing steps highlighted

C. Anshul Bhagtani, Tanupriya Choudhuri (2017) proposes a method of selecting features which have more separable value than others and these features are processed and the performance is evaluated. The proposed algorithm will reduce the testing time as it would check for if the person is suffering from Alzheimer's or not.

Limitations: The proposed algorithm which takes only two values for prediction. By considering just two values they have come to a conclusion that their proposed algorithm achieves higher accuracy when compared to other algorithms

Table -1: Comparison of research work addressing the Prediction of Heart and Alzheimer Disease.

Contributors	Data Mining Algorithm	Accuracy	No. of Attributes used
Heart Disease			
Gouthami Kokkula, 2018[11]	NN-MLP	97%	13
T.Vevikanandan et al., 2017[2]	Fuzzy AHP	98%	9
Theresa Princy. et al., 2016[5]	KNN, ID3	80.62%	5,7
Hlaudi Daniel et al.,2014 [7]	J48, REP TREE, NB	97.22%	11
Alzheimer's Disease			
Tingyan Wang andJason L. Qiu., 2018 [8]	EDM, LSTM	95%	17
Anshul Bhagtani, 2017[9]	NB	91%	20
Veeramuthu et al., 2014[10]	CAD, AR	90%	18

3. CONCLUSIONS

In this paper, survey is conducted on various data mining techniques useful for the prediction and diagnosis of heart disease and Alzheimer disease. The authors have proposed several algorithms but there are certain limitations to these algorithms. In future, it is necessary to develop an algorithm that can overcome these drawbacks by performing accurately using Hybrid Classification algorithm.

REFERENCES

[1] D. Kornack and P. Rakic, "Cell Proliferation without Neurogenesis in Adult Primate Neocortex," Science, vol. 294, Dec. 2001, pp. 2127-2130, doi:10.1126/science.1065467.

[2] Vivekanandan T et al., "Optimal feature selection using a modified differential evolution algorithm and its effectiveness for prediction of heart disease", www.elsevier.com/locate/compbiomed, https://doi.org/10.1016/j.compbiomed, Pages: 125-136 (2017)

[3] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[4] I.H. Witten, E. Frank, M.A. Hall, C.J. Pal, "Data Mining: Practical machine learning tools and techniques," Morgan Kaufmann, 2016.

[5] Theresa Princy R, "Human Heart Disease Prediction System using Data Mining Techniques", International Conference on Circuit, Power and Computing Technologies [ICCPCT], IEEE (2016)

[6] Venkatalakshmi B et al., "Heart Disease Diagnosis using Predictive Data mining", International Journal of Innovative Research in Science, Engineering and Technology [IJIRSET], ISSN (Online): 2319-8753, Volume: 3, Issue: 3, Pages :18731877 (2014)

[7] Hlaudi Daniel Masethe et al., "Prediction of Heart Disease using Classification Algorithms", Proceedings of the World Congress on Engineering and Computer Science [WCECS], Volume: II, ISBN: 978-988-19253-7-4 ISSN: 2078-0958 (Print), ISSN: 2078-0966 (Online) (2014)

[8] Tingyan Wang1, 3,Jason L. Qiu2, Robin G. Qiu3*, Ming Yu1," Early Detection Models for Persons with Probable Alzheimer's Disease with Deep Learning" Department of Industrial Engineering, Tsinghua University, Beijing 100084, China 978-1-5386-1803-5/18/$31.00 IEEE (2018)

[9] Anshul Bhagtani,Tanupriya Choudhury., An Efficient Survey to Detect Alzheimer Disease Using Data Mining Techniques",978-1-5386-1144-9$31.00 IEEE (2017)

[10] Veeramuthu, A., S. Meenakshi, et al. (2014). A New Approach for Alzheimer's Disease Diagnosis by using Association Rule over PET Images. International Journal of Computer Applications 91(9), 9-14.

[11] Aditi Gavhane, Gouthami Kokkula, "Prediction of Heart Disease Using Machine Learning ",Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018) IEEE Conference Record # 42487; IEEE Xplore ISBN:978-1-5386-0965-1.

[12] Kalaiselvi C, "Diagnosing of heart diseases using Average KNearest Neighbor Algorithm of Data Mining", 2016 InternationalConference on Computing for Sustainable Global Development (INDIACom), IEEE, Pages: 3099-3103 (2016)

[13] Sunil Ray, "Learn Naïve Bayes algorithm" and "Decision tree-Simplified", URL: www.analyticsvidhya.com, Retrieved on-10.01.2018.