ISSN: 0975-3583, 0976-2833 VOL 12, ISSUE 04, 2021

The Comparative Study of Heart Disease Prediction using Machine Learning

Mr.J.Nageswara Rao[1], Dr.R. Satya Prasad[2],
[1]Research Scholar Dept.of CSE, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India,
[2] Professor, Dept.of CSE, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India.

Abstract: Nowadays, people are becoming busier and there is no time to check the health. Based on their jobs and other personal issues in life, the sick increases every day. The common and dangerous disease that can occur to the human is heart disease. Approximately 32% of deaths are due to heart diseases. So, early prediction of heart disease may improve the prevention of heart attacks. Machine Learning (ML) algorithms most widely used to process complex data, unstructured data and medical data. Medical data such as disease prediction is most widely used by ML algorithms. ML is a subset of artificial intelligence (AI) which focuses mainly on making predictions based on its experience, analyzing the previous data. In this paper, a comparative study of various machine learning algorithms is discussed and shows the result.

Keywords: ML, heart disease prediction, AI.

1.Introduction

According the World Health Organization (WHO) approximately there are 18.1 million people are died with heart disease in 2017 which are represented 32% of deaths all over the world. Diagnosing and screening of heart diseases becomes more complicated to predict or to analyze with manual process. Machine Learning (ML) is considered as most suitable and compatible algorithms for detecting the heart diseases. Doctors and experts are working on ML algorithms to get the accurate results. ML algorithms are more powerful in pattern recognition and classification which is compared with the other existing algorithms. In middle-income countries huge number of deaths caused by the heart diseases [12,3,4]. Many factors shows impact on humans based on their life style, personal and professional habits that affects the heart disease. To prevent the deaths from the heart disease an effective and intelligent system have to be developed to predict the heart disease in the early stages. Data mining (DM) is used to extract the needful information from the large datasets that can used in several fields such as medical,

ISSN: 0975-3583, 0976-2833 VOL 12, ISSUE 04, 2021

education and business fields. ML is very fastly involves the domains of artificial intelligence. ML algorithms show the more attention on medical field. ML is the best replacement for regular prediction models that gains more accuracy [5].

Indeed, ML is a thought which allows the machine to realize from models and knowledge, which too without being expressly customized. So instead of you composing the code, what you are doing is you feed information to the traditional calculation, and therefore the calculation/machine fabricates the rationale logic in to the given information.

2.Literature Survey

Heart plays the major role in human body which pumps the blood to all the body. A huge protection is required to the human heart. Several ML algorithms are applied on various datasets and show the performance of ML algorithms. Based on the dataset complexity the performance of algorithms are analyzed [6]. The author [7] proposed the ML algorithm called as naive bayes (NB) that increase the performance in terms of low variance and high accuracy when this is compared with the K-Nearest Neighbor (KNN). KNN has the main drawback such as over fitting, with this the performance is reduced. One more disadvantage of KNN is this cannot work efficiently on huge datasets.

The author [1,2,3] proposed the new merger approaches that combine the two different approaches as the single algorithm which is called as Ensemble algorithm. The comparison between the several algorithms such as KNN, decision tree (DT), NB and hybrid approach is applied on heart datasets. The overall accuracy is 89.1% for the proposed system. To predict the heart disease in early stages the author [6] proposed the new interesting pattern and knowledge that are extracted from the huge datasets. After the comparison between several ML algorithms the SVM shows the better performance with 93.2% accuracy.

The author introduced the enhanced algorithm which is merged with SVM gets the huge accuracy with 96.7%. The proposed algorithm is applied on UCI repository dataset with 310 samples with 15 features and obtained the better performance and comparatively it is good among other algorithms. The author [8] worked on various multi layered perceptron approach that predicts the heart disease with the increase of high accuracy and the algorithm is adopted

ISSN: 0975-3583, 0976-2833 VOL 12, ISSUE 04, 2021

with CAD technology. If many persons are uses these technology to detect the heart disease then the people get more awareness about this disease and many people can use this prediction to prevent the heart disease.

3.Heart Disease Prediction

Heart disease has made plenty of genuine worries among analysts; one among the many difficulties in coronary illness is the true discovery and finding the presence of it inside a person's. Early methods haven't been such an excellent amount of productive in thinking that even clinical teachers aren't all that effective enough in anticipating coronary illness. There are different clinical instruments accessible on the lookout for foreseeing coronary illness there are two significant issues with them, the primary is that they're tons costly, and therefore the subsequent one is that they're not effectively able to compute the chance of coronary illness in a human.

MI is one such instrument that is broadly utilized in various spaces since it doesn't need a diverse calculation for the distinctive dataset. Reprogrammable limits of MI bring plenty of solidarity and open new entryways of chances for zones like clinical science.

An Enhanced Novel Dynamic Data processing (ENDDP) algorithm

ENDDP is a statistical classifier which assumes no suppression between attributes. Some of the bayes features are integrated into this to analyze the attributes that are independent of each other. The working process of the proposed classifier is as follows:

Training Step: Training plays the major role in applications such as heart disease prediction. This is the proposed training which selects the dataset data randomly to perform well. This method expects the parameters of a probability distribution known as the prior probability from the training data.

Deep Boltzmann Machine Algorithm

DBM is the deep learning algorithm which follows the multilayer structure integrated with the undirected graph. This is very different from traditional restricted Boltzmann machine (RBM). The DBM process with the multiple layers and RBM is consists of only one layer. Several

ISSN: 0975-3583, 0976-2833 VOL 12, ISSUE 04, 2021

experiments shows that DBM performed well when compare with RBM. DBM can work efficiently on heart disease prediction. This also called as three layer creative model. In the last layers, the bi-directional layers are used.

$$E(v,h) = -\sum_{i}^{i} v_i b_i - \sum_{n=1}^{N}\sum_{k} h_{n,k} b_{n,k} - \sum_{i,k} v_i w_{ik} h_k - \sum_{n=1}^{N-1}\sum_{k,l} h_{n,k} w_{n,k,l} h_{n+1,l}$$

DBM consists of 'N' hidden layers.

In this paper, initialize the two layers Boltzmann machine within –layer connections. The final state is $\{v, h^1, h^2\}$ is defined as:

$$E(\mathrm{v},\mathrm{h}^1,\mathrm{h}^2;\theta) = -\mathrm{v}^{\mathrm{T}}\mathrm{W}^1\mathrm{h}^{1\mathrm{T}}\mathrm{W}^2\mathrm{h}^2$$

Where $\theta = \{\mathrm{W}^1, \mathrm{W}^2\}$ are the model parameters, these represents the visible-to-hidden and hidden-to-hidden. The probability that the model gives for the disease prediction is in visible vector v is:

$$p(v;\ \theta) = \frac{1}{Z(\theta)} \sum_{h^1 h^2} \exp(\mathrm{E}(\mathrm{v},\mathrm{h}^1,\mathrm{h}^2;\theta))$$

Minimum Redundancy Maximum Relevance (MRMR)

The feature selection algorithm is used to identify the characteristics of medical datasets especially heart disease and minimize their relevance accordingly explained with its pairing with the relevant feature selection algorithm. The process of this algorithm based on features that are mutually far away and having a "high" correlation to the variable in classification. In ML, feature extraction is an important sub-domain that chooses a subset of data that are equal to a specific problem domain and is known as Maximum Relevance. The subset consists of redundant entries and MRMR aims to remove those redundant entries. MRMR has a different type of application such as speech recognition and Cancer Diagnosis. Using this algorithm we can extract features in many ways. This technique is known as "Minimum Redundancy Maximum Relevance" and is found those feature subsets are more influential than the Maximum relevance technique. In this stage, the feature extraction algorithm MSMR works to increase the province between the reciprocal distribution of selected features and the classification variable. The main objective of

ISSN: 0975-3583, 0976-2833 VOL 12, ISSUE 04, 2021

this algorithm is to select features by using mutual information, correlation, or distance score. Resulted features are mutually correlated and cover a narrow region in space. The feature set which is the relevance initialize with 'S' for class 'c' is explained by the average of all reciprocal data values between individual feature ‘fi’ and the class ‘c’ is defined by the given equation:

$$D(S,c) = \frac{1}{|S|}\sum_{fi\in S} I(fi,c)$$

'S' represents the redundancy of all features is the average of all the reciprocal data between features ‘fi’ and ‘fj’ is shown below and described by the given equation

$$R(S) = \frac{1}{|S|^2}\sum_{fi,fj\in S} I(fi,fj)$$

MRMR is the combination of above and below equations and is defined as:

$$MRMR = \max_S \left[\frac{1}{|S|}\sum_{fi\in S} I(fi,c) - \frac{1}{|S|^2}\sum_{fi,fj\in S} I(fi,fj)\right]$$

Dataset Description

The experiments are conducted on "Heart datasets which is called as UCI Cleveland dataset collected from the UCI website with 75 attributes and 15 features [1,2,3,8]. This dataset having 350 samples of several age patients that are affected with heart disease and abnormal patients. 210 samples are belongs to male and 140 are female samples. 190 samples are with major health issues and 160 samples are with normal sand simple health issues. The algorithm works on predicting the heart disease from the 350 samples.

4.Results

Data Mining Techniques	Accuracy	Sensitivity	Specificity
Naive Bayes	91.42%	86.43%	79.76%
Random Forest	89.56%	89.76%	84.54%
Decision Tree	92.32	87.12	86.67
ENDDP	97.98%	97.45%	98.54%
EDDA	98.12%	98.45%	99.12%
EFEDLA	98.98%	99.10%	99.60%

ISSN: 0975-3583, 0976-2833 VOL 12, ISSUE 04, 2021

Table: 1 comparative result

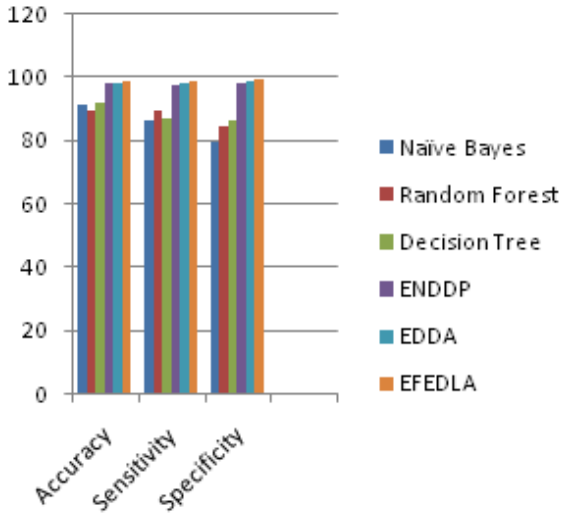

Figure: 1 the performance of the various machine learning and proposed methodology

5.Conclusion

In this paper, various proposed and existing methodologies are developed to predict the heart disease accurately. Machine Learning is most widely fit for processing the heart disease prediction because only ML algorithms can compatible to overcome the issues in heart disease prediction.

References

References

[1]. J. Nageswara Rao, Dr. R. Satya Prasad, "An Enhanced Novel Dynamic Data Processing (ENDDP) Algorithm for Predicting Heart Disease in Machine Learning", IJSRCSEIT, ISSN : 2456-3307, Volume 7 Issue 1, pp. 94-104, January-February 2021.

[2]. J. Nageswara Rao, Dr. R. Satya Prasad, "An Ensemble Deep Dynamic Algorithm (EDDA) to Predict the Heart Disease", IJSRSET, Online ISSN: 2394-4099, Print ISSN : 2395-1990, Volume 8 Issue 1, pp. 105-111, January-February 2021.

[3].J. N. Rao and M. Ramesh, "A Review on Data Mining & Big Data Machine Learning Techniques", JRTE, vol. 7, no. 6S2, pp. 914-916, April 2019.

ISSN: 0975-3583, 0976-2833 VOL 12, ISSUE 04, 2021

[4] Seckeler MD, Hoke TR. The worldwide epidemiology of acute rheumatic fever and rheumatic heart disease. Clin Epidemiol. 2011;3:67.

[5] Gaziano TA, Bitton A, Anand S, Abrahams-Gessel S, Murphy A. Growing epidemic of coronary heart disease in low-and middle-income countries. Curr Probl Cardiol. 2010;35(2):72–115.

[6] Aditi Gavhane, Gouthami Kokkula, Isha Panday, Prof. Kailash Devadkar, "Prediction of Heart Disease using Machine Learning", Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology(ICECA), 2018.

[7] Himanshu Sharma and M A Rizvi, "Prediction of Heart Disease using Machine Learning Algorithms: A Survey" International Journal on Recent and Innovation Trends in Computing and Communication Volume: 5 Issue: 8 , IJRITCC August 2017.

[8] Senthil kumar mohan, chandrasegar thirumalai and Gautam Srivastva, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques" IEEE Access 2019.