ISSN: 2229-6956 (ONLINE)
DOI: 10.21917/ijsc.2021.0332

AN EFFECTIVE HEART DISEASE PREDICTION USING MACHINE LEARNING TECHNIQUE

D. Komalavalli, R. Sangeethapriya, R. Indhu, N. Kanimozhi and G. Kasthuri
Department of Information Technology, Sona College of Technology, India

Abstract

Heart disease is the foremost significant causes of transience within the world nowadays. It is a vital challenge to predict the cardiovascular disease in the range of clinical data investigation. Machine Learning (ML) is the most popular and powerful approach that has been appeared to be effective in making decisions and predictions from the huge amount of information delivered by the healthcare industry. ML techniques are also used in recent developments in wide areas of the Internet of Things (IoT). There are various studies done to predict the heart disease with ML techniques and it gives only a glimpse of it. In this paper, a simple TensorFlow model is proposed to find out major features by applying ML techniques that result in better accuracy in the prediction of cardiovascular disease. The prediction model is presented with diverse combinations of features and known classification algorithms. This version for coronary heart disorder with the ML based TensorFlow Model produces a more desirable overall performance with a higher accuracy stage in prediction.

Keywords:

Machine learning, heart disease prediction, feature selection, binary classification, TensorFlow

1. INTRODUCTION

Heart Diseases describe a range of conditions that affect the heart and stand as a leading cause of death all over the world [2]. Due to heart disease there are several million people are dying in each year. It is a severe disease that affects the main function of the heart. It is important for many facets of health promotions and clinical medicines that to estimate the risk of coronary heart diseases of people. With the rapid growth and development of digital technologies in health care centers for storing huge amount of medical data in their databases and those are very complex and challenging to analysis in nature. There are many popular techniques available and plays vital roles in analyzing of different data in medical databases. These techniques and algorithms can be directly applied on a dataset to create models and inferences from the dataset. The prediction of heart disease is one of the challenging task and risk factor in the medical industry and there is a need in medical domain application to detect the heart diseases and its related symptoms through computer-aided diagnosis methods and applications, in which the data can be collected from many sources and are evaluated based on computer based applications. There are several methods like clustering, classification, Decision Support Systems (DSS), Genetic Algorithm (GA), Neural Network (NN) and many more to find out the austerity of the patients clinical data. A fuzzy rule based [1], [14] system can be used for predicting the heart disease. The clinical dataset may be collected from UCI repository. The data has to be properly cleaned in the pre-processing stage. The both training and test data are classified based on the risk level. An automated weighted fuzzy rule based system is to analyse the risk level and classify the relevant attributes to find out the deviations in it. The decision rules are generated to find the frequency from the deviations occurred. Finally fuzzy rules are applied to find the risk prediction in the clinical dataset. The performance of the fuzzy ruled based prediction system is better in compared with the Neural Network system. A Multilayer Perception Algorithm (MLP) [2] is a popular method that gives an efficient training with the optimization technique like gradient descent. It is one of the NN models that has same architecture of Feed-Forward back propagation for supervised training. The heart disease dataset is classified [5] using one of the data mining technique called GA classifier. It uses the association rule mining to classify and predict the symptoms of heart diseases with the Cleveland dataset which is collected from UCIML repository. The accuracy, Sensitivity and Specificity are obtained and compared with methods like NN and Naive Bayes algorithms. This system gives the good results in the prediction of heart disease but the better results can be obtained in combining this GA classifier with other newer Machine Learning (ML) algorithms using hybrid techniques. The hybrid model [11] with less input variables are given the good classification results and in improved accuracy of results. Since most of the studies uses all the 13 variables in the UCIML repository dataset to classify the heart disease. Logistic Regression (LR), Multivariate Adaptive Regression Splines (MARS), Artificial Neural Network (ANN) and Rough set techniques are used in this model for prediction. The hybrid method [12] with Random Forest (RF) and Linear Model are used to predict the cardiovascular diseases (CVD) and produced an enhanced performance level with the best accuracy of results in the prediction model for heart disease. Hence, there is a need of invention in newer concepts to improve the prediction of heart diseases with the highest accuracy in a short span of time. In this work, a simple ML technique is used to predict the heart disease with UCIML repository datasets using Tensor Flow analysis.

2. RELATED WORK

A weighted fuzzy rule based Clinical Decision Support System (CDSS) [1] is presented to analysis the heart disease. It is automatically collects the information from the patient's clinical dataset obtained from UCI repository using the fuzzy rules. As a first step, data pre-processing is applied to remove the missing values and noise from the clinical dataset. Then the fuzzy system is used to define the weighted fuzzy rules and feature extraction. The decision rules are constructed and are applied to the database to find the frequency model. Based on the frequency level, the weight is calculated for every decision rule obtained from the training dataset and weighted fuzzy rules are obtained with the help of fuzzy membership function. Then these inputs are given to the proposed system to classify and predict the heart diseases.

A back propagation MLP [2], [4], [8] algorithm is proposed to predict the heart diseases. It computes the depth of loss function in the input data which includes all the weights in the network.

Then, the gradient techniques are used to optimize the weights for minimizing the loss function in the network. In general, the MLP algorithm has an input, one or more hidden layers and an output layer. It represent any input-output relationship with a finite number of discontinues and sufficient neurons in the hidden layers. Hence, it gives more accuracy of results in predicting hear diseases with clinical datasets from UCI repository.

Classification models [3], [15] of data mining are analyzed with Decision Trees, Neural Networks and Naive Bayes Classifier to predict the heart disease. GA classifier [5] is presented based on association rule mining [7] method to classify the clinical dataset and to predict the heart disease and to detect sick and healthy conditions of both males and females. It uses the Cleveland dataset from UCIML repository to analyze the patient data. The initial rules are represented in Chromosome format then effective rules are extracted to measure the classifier performance using the test dataset on the inferred rules. Finally, the performance is measured based on the parameters like accuracy, sensitivity and specificity.

Computational Intelligence method [6] is proposed to analyze the responsible risk factors for heart disease with the four performance metrics of accuracy, True Positive rate (TP), F-Measure and time. A train-test split on the dataset and then performed the 10-fold cross validation to select the best performer for training. An ANN method [10] is used to train and accurately analyze the clinical and ECG dataset and diagnose the heart to predict the abnormities in an efficient manner. First, it acquire the ECG data of the patients in analog format. Then this analog data format is converted into digital format using A/D converter and finally Neural Network algorithm is used to classify the heart disease and to predict the abnormalities in the heart or it's functioning.

ID3 algorithm [16] is implemented to predict the similarities of disease that are identified through coronary illness. This paper presents a programmed and obscured way to deal with coronary diseases through the framework and it also helps in reduce the death rate of people by this diseases.

K-Star algorithm [17], [18] is presented predicting coronary heart diseases. In this work an expectation framework for heart infection in utilizing the Learning Vector Quantization neural system calculation is exhibited. The neural system in this framework recognizes 13 clinical datasets information and predicts that there is a nearness of coronary illnesses in the patient.

Vertical system integration [19] of a sensor node and toolkit of machine learning algorithm is presented for predicting the heart disease of a person. The dataset used in this model is automatically taken from the raw value of the heart pulse sensor and it is also used some manually given data for the prediction.

Supervised machine learning techniques [20], [21] using Naïve Bayes, decision tree, random forest and KNN based research was done on UCI repository to predict whether a patient affected through heart disease or not. Various experiments were conducted using different classifier algorithms through WEKA tool.

Hybrid models [11], [12] are proposed to and analyze the clinical dataset and to predict the heart disease in the earlier stage to save the patient life. Many ML algorithms are combined to get the good and efficient results from the clinical datasets. Minimum input variables are used in the input dataset for prediction. It produced good results instead of taking out all the parameters in the UCI dataset.

3. METHODOLOGY AND SYSTEM DESIGN

In this methodology, as a first step patient data is collected from UCI dataset repository [13] and data pre-processing is done with the clinical dataset to remove the missing values and noisy information. After this process, a Neural Network model is created using binary classifier in the TensorFlow. TensorFlow is an open source machine learning framework to implement applications and to develop and research using machine learning and deep learning applications. Our model is then trained to predict the heart disease from patient data. The Fig.1 illustrates the complete methodology of the system.

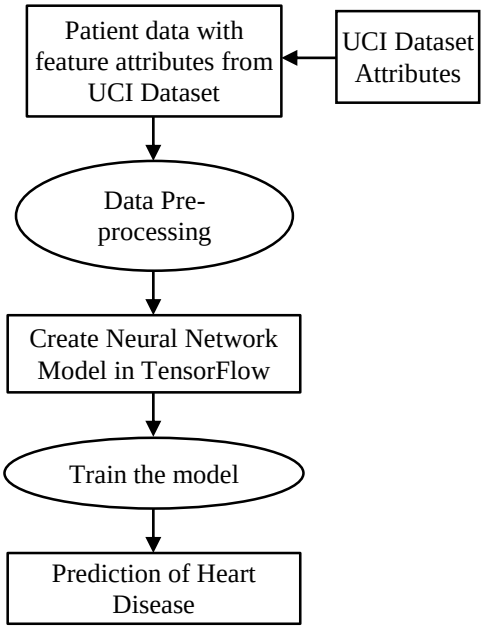

Fig.1. System Design Architecture

4. PROPOSED METHOD

The proposed method includes different algorithms of data mining and neural network models. The dataset is retrieved from UCIML repository. The steps include data collection, data pre-processing, feature extraction, training and prediction. Data collection and preprocessing steps are performed using data mining techniques. Then data is converted into TensorFlow dataset. This is given to the system for splitting out the data into train sets and test sets. A Neural Network based binary classifier model is used to train the model to obtain the loss function. Finally, the prediction of heart disease is done based on the accuracy of the test set.

Algorithm 1: Heart Disease Prediction using TensorFlow

Step 1: Load the patient ECG data from UCI dataset repository.

Step 2: Apply the data pre-processing technique to remove the missing and noisy information

Step 3: Convert the dataset into TensorFlow dataset.

Step 4: Use binary classifier model to split the data into train set and test set.

Step 5: Train the model using loss function obtained from binary function

Step 6: Loss function is binary cross-entropy defined by

$$-(y \log(p)+(1-y)\log(1-p) \quad (1)$$

where y is binary function.

Step 7: Predict the heart disease based on the accuracy of test sets.

The Table.1 shows sample of training process generated by TensorFlow model. Accuracy is recorded in each of the training set to do the prediction in the better manner.

Table.1. Sample of training process

Iteration	Accuracy	Loss	Validation Accuracy	Validation Loss
95	0.3018	0.8430	0.4012	0.8689
96	0.2882	0.8547	0.3436	0.8689
97	0.2889	0.8732	0.3368	0.8689
98	0.2964	0.8386	0.3537	0.8770
99	0.3062	0.8282	0.4110	0.8607
100	0.2685	0.8821	0.3669	0.885

4.1 LIST OF ATTRIBUTES IN UCI DATASET

The Table.2 depicts the description and type of attributes from UCI dataset [13]. There are 14 attributes that feature in the prediction of heart disease, where only one attribute serves as the output or the predicted attribute to the presence of heart disease in a patient.

Table.2. UCI Dataset Attributes

Attributes	Description	Type
Age	Patient's age in completed years	Numeric
Sex	Patient's Gender	Nominal
Cp	Type of Chest pain	Nominal
Trestbps	Level of blood pressure	Numeric
Chol	Serum cholesterol in mg/dl	Numeric
FBS	Blood Sugar level	Nominal
Resting	Results of electrocardiogram	Nominal
Thali	Maximum rate of Heart	Numeric
Exang	Angins induced by exercise	Nominal
OldPeak	Exercise-induced ST depression	Numeric
Slope	ST segment measured	Nominal
Ca	Fluoroscopy coloured major vessels numbered from 0 to 3	Numeric
Thal	Status of the heart	Nominal
Num	Heart disease diagnosis	Nominal

5. EVALUATION RESULTS

The prediction models are developed using 14 features and the accuracy is calculated for modeling techniques.

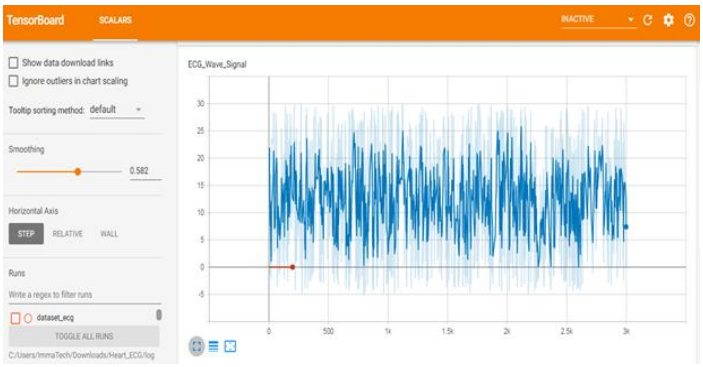

Fig.2. Normal ECG Dataset

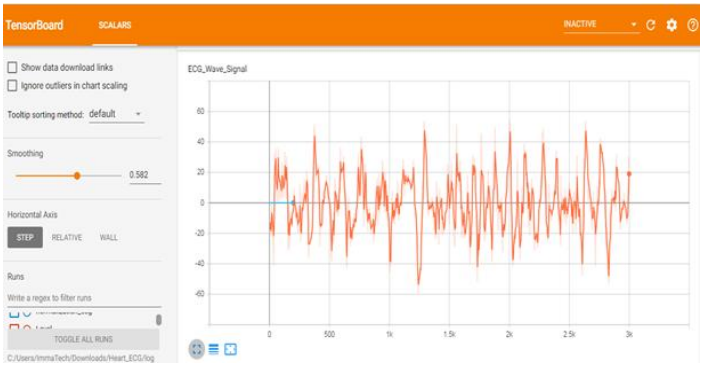

Fig.3. User ECG Dataset

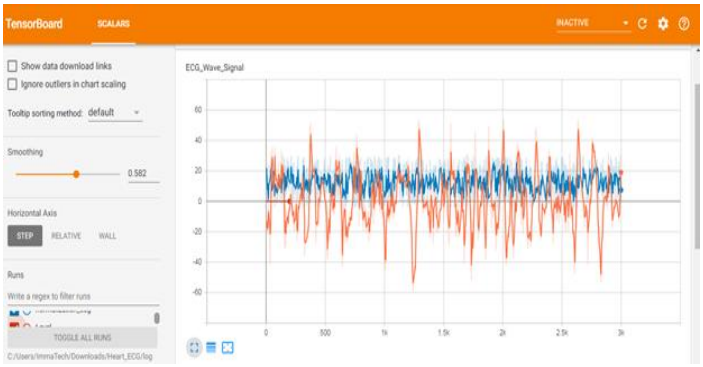

Fig.4. Analyzing heart disease with ECG Comparison

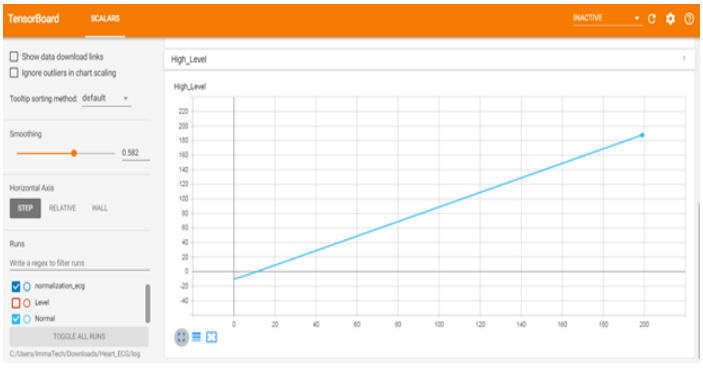

Fig.5. Normal stages

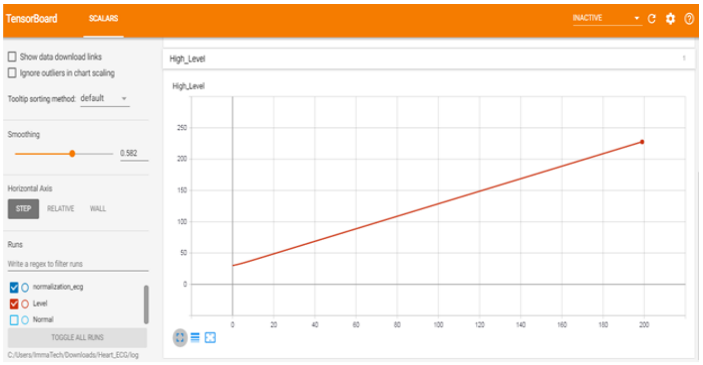

Fig.6. User Level

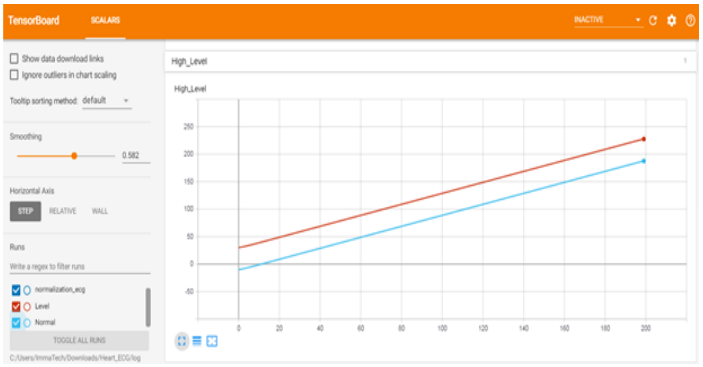

Fig.7. Comparison of Heart disease level

These output is obtained from the TensorFlow model that shows the comparison of normal ECG and User ECG pulse rate from Fig.2 to Fig.4. The highest accuracy is obtained from this model in predicting the heart disease with normal ECG rate in comparison with user ECG pulse rate.

6. CONCLUSION AND FUTURE WORK

Identifying the healthcare data processing of heart will help in long term saving of human lives and early detection of abnormalities in heart conditions. Machine learning techniques were used in this work to process raw data and provide a new and novel discernment towards heart disease. Heart disease prediction is challenging and very important in the medical field. However, the mortality rate can be drastically controlled if the disease is detected at the early stages and preventative measures are adopted as soon as possible. Further extension of this study is highly desirable to direct the investigations to real-world datasets instead of just theoretical approaches and simulations. The proposed Neural Network based TensorFlow model proved to be quite accurate in the prediction of heart disease.

The future work of this research can be performed with diverse mixtures of machine learning techniques to better prediction techniques. Furthermore, new feature selection methods can be developed to get a broader perception of the significant features to increase the performance of heart disease prediction.

REFERENCES

[1] P.K. Anooj, "Clinical Decision Support System: Risk Level Prediction of Heart Disease using Weighted Fuzzy Rules", *Journal of King Saud University - Computer and Information Sciences*, Vol. 24, No. 1, pp. 27-40, 2012.

[2] M. Durairaj and V. Revathi, "Prediction of Heart Disease using Back Propagation MLP Algorithm", *International Journal of Scientific and Technology*, Vol. 4, No. 8, pp. 235-239, 2015

[3] M. Gandhi, "Predictions in Heart Disease using Techniques of Data Mining", *Proceedings of International Conference on Futuristic Trends on Computational Analysis and Knowledge Management*, pp.520-525, 2015.

[4] A. Gavhane, "Prediction of Heart Disease using Machine Learning", *Proceedings of International Conference on Electronics, Communication and Aerospace Technology*, pp.1275-1278, 2018.

[5] P.S. Kumar, "A Computational Intelligence Method for Effective Diagnosis of Heart Disease using Genetic Algorithm", *International Journal of Bio-Science and Bio-Technology*, Vol. 8, No. 2, pp. 363-372, 2016.

[6] J. Nahar and T. Imam, "Computational Intelligence for Heart Disease Diagnosis: A Medical Knowledge Driven Approach", *Expert Systems with Applications*, Vol. 40, No. 1, pp. 96-104, 2013.

[7] J. Nahar and T. Imam, "Association Rule Mining to Detect Factors which Contribute to Heart Disease in Males and Females", *Expert Systems with Applications*, Vol. 40, No. 4, pp. 1086-1093, 2013.

[8] Tulay Karaylan and Ozkan Kilic, "Prediction of Heart Disease using Neural Network", *Proceedings of International Conference on Computer Science and Engineering*, pp. 719-723, 2017.

[9] T.P. Thomas, "Human Heart Disease Prediction System using Data Mining Techniques", *Proceedings of International Conference on Circuit, Power and Computing Technologies*, pp. 12-18, 2016.

[10] D.K. Ravish and N.R. Shenoy, "Heart Function Monitoring, Prediction and Prevention of Heart Attacks: using Artificial Neural Networks", *Proceedings of International Conference on Contemporary Computing and Informatics*, pp. 1-6, 2014.

[11] Y.E. Shao, C. Hou and C. Chiu, "Hybrid Intelligent Modeling Schemes for Heart Disease Classification", *Applied Soft Computing*, Vol. 14, No. 1, pp.47-52, 2014.

[12] Senthilkumar Mohan, "Effective Heart Disease Prediction using hybrid Machine Learning Techniques", *IEEE Access*, Vol. 7, pp. 81542-81554, 2016.

[13] C.L. Blake and C.J. Merz, "UCI Repository of Machine Learning Databases", Available at http://www. ics. uci. edu/~ mlearn/MLRepository. html, Accessed at 1998.

[14] P. Shanmugaraja, K. Chokkanathan, J. Anitha, A. Parveen Begam and N. Naveenkumar, "Dynamic Packet Scheduler for Queuing Real Time and Non-Real Time Internet Traffic", *International Journal of Recent Technology and Engineering*, Vol. 8, No. 3, pp. 1-12, 2019.

[15] J. Akilandeswari and G. Jothi, "Performance Comparison of Machine Learning Algorithms that Predicts Students' Employability", *Proceedings of International Conference on Information and Communication Systems*, pp. 221-228, 2017.

[16] L. Sathish Kumar and A. Padmapriya, "Prediction for Common Disease using ID3 Algorithm in Mobile Phone and

Television", *International Journal of Computer Applications*, Vol. 50, No. 4, pp. 23-29, 2012.

[17] Wiharto Wiharto and Hari Kusnanto, "Intelligent Systems for Diagnosis Level of Coronary Heart Disease with K-Star Algorithm", *Healthcare Informatics Research*, Vol. 22, No. 1, pp. 1-15, 2016.

[18] S. Jayashril and D.R.P. Sonawane, "Prediction of Heart Disease using Learning Vector Quantization Algorithm", *Proceedings of International Conference on Business, Industry and Government*, pp. 451-456, 2014.

[19] S. Nandhini and Monojit Debnath, "Heart Disease Prediction using Machine Learning", *International Journal of Recent Engineering and Research and Development*, Vol. 3, No. 10, pp. 39-46, 2018.

[20] Devansh Shah and Samir Patel, "Heart Disease prediction using Machine Learning Techniques", *SN Computer Science*, Vol. 1, No. 6, pp. 1-13, 2020.