ACTIVE PREDICTION OF CARDIOVASCULAR DISEASE USING HYBRID LEARNING STRATEGIES

[1]Rachana T, [2]Gowthami H R, [3]Ramakrishna M, [4]Apoorva R

ALVA'S INSTITUTE OF ENGINEERING AND TECHNOLOGY
INFORMATION SCIENCE AND ENGINEERING, Mangalore, India

***Abstract*: Heart disease is one of the most important causes of death in the world today, Prediction for cardiovascular disease is a key problem in the world of clinical data analysis Machine learning (ML) has been shown to be effective in helping to make decisions and predictions based on the large amount of data produced by the healthcare industry. We also saw the use of ML techniques Heart disease is one of the most important causes of death in the world today, Prediction for cardiovascular disease is a key problem in the world of clinical data analysis machine learning(ML) has been shown to be effective in helping to make decisions and predictions based on the large amount of data produced by the healthcare industry, We also saw the use of ML techniques used in recent developments in different areas of the Internet of Things (IoT), Various studies provide only insight into detecting heart disease using ML techniques.**

***Index Terms*: Decision tree, KNN, Machine learning.**

I. INTRODUCTION

Due to several common risk factors such as diabetes, high blood pressure, high cholesterol, irregular pulse rate and many other causes, it is difficult to identify heart disease. Different techniques have been used in data mining and neural networks to determine the truth of cardiac disease among humans. Based on various approaches such as K-Nearest Neighbor Algorithm (KNN), Decision Trees (DT), Genetic Algorithm (GA), and Naive Bayes(NB), the extent of the infection is graded. The essence of heart disease is complex and it is therefore necessary to treat the condition carefully. It can damage the heart and cause premature death if not done so. The medical science and data mining approach is used to diagnose different types of metabolic syndromes. Classified data mining plays an important role in the forecasting of heart disease and data analysis.

We also saw the use of decision trees to forecast the precision of heart disease incidents various methods were used to interpret information by using known data mining tools to predict heart disease, Numerous readings were performed in this study to generate a prediction model using not only distinct techniques but also two or more techniques. These mixed new techniques are commonly referred to as blended processes, Using heart rate time series, they implement neural networks. This method uses various forecasting patient reports such as Left Bundle Branch Block (LBBB), Right Bundle Branch Block (RBBB), Atrial Fibrillation(AFIB),Normal Sinus Rhythm (NSR), Sinus Bradycardia (SBR), Atrial Flutter(AFL),Premature Ventricular Contraction(PVC)),and Second Degree Block (BII) to assess the precise heart disease state of the patient. The Radial Basis Function Network (RBFN) dataset is used for classification, where 70% of the data is used for learning and the other 30% is used for classification.

II. RELATED WORK

In the fields directly related to this paper, there is ample related research. To generate the highest prediction of accuracy in the medical field, ANN was launched. ANN's multilayer interpretation of the back propagation (MLP) is used to predict heart disease. The results obtained were compared and found to be better with the results of existing models within the same field. UCI laboratory data collected from patients with heart disease is used to identify patterns with NN, DT, SVM Support Vector machines, and Naive Bayes. The results are compared for performance and accuracy with these algorithms. The hybrid approach suggested yields results for F-measure of 86.8 percent, compared with other existing methods. Convolutionary Neural Networks (CNN) classification without segmentation is added. This approach considers in the training phase the heart cycles with different starting locations from the signals of the electrocardiogram (ECG). CNN can create features with different positions in the patient's test phase. A large amount of data produced by the medical industry has not previously been used effectively.
The new approaches discussed here minimize costs and improve cardiac forecasting in an easy and effective manner. The different research approaches considered in this work for the prediction and classification of heart disease using techniques of ML and deep learning (DL) were highly accurate in assessing the effectiveness of these methods.

III. LITERATURE SURVEY

1. Title: A data mining model for predicting the coronary heart disease using random forest classifier
Author: A. S. Abdullah and R. R. Rajalaxmi,

Coronary heart disease (CHD) is a common form of heart disease and a major cause of premature death. From the medical science point of view, data mining is involved in the identification of different types of metabolic syndromes. In forecasting and data discovery, classification techniques in data mining play an important role. Technique of classification such as Decision Trees A

data mining framework was developed in this paper using the Random Forest classifier to improve predictive reliability and investigate various CHD-related events. The model will assist medical practitioners with their different events to assess CHD and how it could be applied to specific segments of the population. The examined cases were Angina, AMI (Acute Myocardial Infarction).

2. Title: Using PSO algorithm for producing best rules in diagnosis of heart disease
Author: A. H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori, and M. Abdar.

Heart disease continues to be an increasing global health issue. Limiting human experience and expertise in manual diagnosis in the health care system leads to inaccurate diagnosis, and information about various diseases is either inadequate or inaccurate as it is gathered from different types of medical equipment. Since a person's condition is correctly predicted, it is very important Doctors can reduce mistakes and financial losses by equipping medical science with intelligent instruments for the diagnosis and treatment of disease. In this paper, the Particle Swarm Optimization (PSO) algorithm is used to produce rules for heart disease, which is one of the most efficient evolutionary algorithms. The random rules are first encoded and then optimized using their accuracy using pso algorithm.

3. Title: Back propagation neural network for prediction of heart disease
Author: N. Al-milli

Recently, the researchers have suggested many applications, methods and various algorithms to build successful medical decision support systems. In addition, the creation and representation of new algorithms and new methods continues day by day. Diagnosing heart disease is one of the key issues and many researchers have been investigating the development of intelligent medical decision support systems to improve physicians ' ability. Neural network is a tool widely used to predict the diagnosis of heart disease. A prediction system for heart disease is developed using neural network in this research paper. The proposed system used 13 heart disease prediction medical attributes. The experiments conducted in this study demonstrated the proposed algorithm's good performance relative to similar state-of - the-art approaches.

4. Title: Analysis of neural networks based heart disease prediction system
Author: C. A. Devi, S. P. Rajamhoana, K. Umamaheswari, R. Kiruba, K. Karunya, and R. Deepika

Heart disease is one of the main causes of higher death rates. Healthcare is one of the main winners of information and analytics. Extracting medical data is increasingly required to predict and treat high death rates due to heart attack. Data terabytes are generated daily. To prevent poor medical decisions leading to disastrous consequences, quality services are needed. Hospitals can use appropriate decision support systems to reduce the expense of clinical testing.
Hospitals nowadays use information systems from hospitals to manage patient data. There is no efficient use of huge amounts of data produced by the healthcare industry. To that the cost and predict heart disease in an easy way, some new approach is needed. The purpose of this paper is to review various works of research on prediction and classification of heart disease using different techniques of machine learning and deep learning and to conclude that techniques are efficient and accurate.

5. Title: Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules
Author: P. K. Anooj

As people have recently had an interest in their health, one of the most active research areas was the development of medical domain application. One example of the clinical field application is the heart disease detection system based on computer-aided methods of diagnosis, where information is collected from some other sources and analyzed based on computer-based applications. Previously, software use was to develop a knowledge-based clinical decision-making support system that manually transferred this information to computer algorithms using expertise from medical experts. This process takes time and relies on the views of medical experts that may be subjective. Machine learning techniques have been developed to automatically gain knowledge from examples or raw data to handle this problem.
For the diagnosis of heart disease, a weighted fuzzy rule-based clinical decision support system (CDSS) is provided, automatically gaining information from the clinical data of the patient. The proposed framework of clinical decision aid for cardiac risk prediction of heart in two phase (1) an automated approach to generating weighted fuzzy rules and (2) a fuzzy rule-based decision-making support system.
In the first step, in order to obtain the weighted fuzzy rules, we used the mining technique, attribute selection and attribute weight age process. Then, the fuzzy system is constructed in accordance with the weighted fuzzy rules and chosen attributes. Instead, according to the weighted fuzzy rules and selected attributes, the fuzzy structure is constructed. Finally, the testing is carried out on the proposed system using the data sets collected from the UCI database and the system output is compared.

IV. PROPOSED METHODOLOGY

We used a R studio rattle in this study to perform the Cleveland UCI repository classification of heart disease. It provides an easy-to-use visual representation of the dataset, work environment and predictive analytics construction.ML begins with a pre-processing data phase followed by feature selection based on DT entropy, modeling performance evaluation classification, and results with improved accuracy. Selection and simulation of the function will continue to be repeated for different combinations of attributes.

A. DATA PRE-PROCESSING

Information on heart disease is pre-processed after different documents have been compiled. A maximum of 303 patient records are included in the database, where 6 records are with some missing values. These six records are excluded from the registry and the remaining 297 records of patients are used in pre-processing. For the attributes of the specified data set, the multi-class parameter and binary classification are implemented. T Data pre-processing is done by converting medical records into values of diagnosis. The findings of pre-processing data for 297 patient records suggest that 137 records show the value of 1 for the presence of heart disease, while the remaining 160 displayed the value of 0 for the absence of heart disease.

B. FEATURE SELECTION AND REDUCTION

Two age and gender attributes are used to classify the patient's personal information from among the 13 attributes of the data set. Because they contain vital clinical records, the remaining 11 attributes are considered important.
In order to diagnose and learn the severity of heart disease, clinical records are vital. Several (ML) techniques are used, namely NB, GLM, LR, DL, DT, RF, GBT and SVM, as described in this experiment.

Algorithm 1 Decision tree based partition

Requirement: Input: D dataset –features with a target class
for features do

for Each sample **do**
Execute the Decision Tree algorithm **end for**
Identify the feature space f_1, f_2, ..., f_x of dataset UCI.

end for
Obtain the total number of leaf nodes l_1, l_2, l_3,..., l_n with its constraints
Split the dataset D into d_1, d_2, d_3,..., d_n based on the leaf nodes constraints.
Output: Partition datasets d_1, d_2, d_3,..., d_n

Algorithm 2 Apply ML to Find Less Error Rate

Require: Input: Datasets with partition – d_1, d_2, d_3,..., d_n
for ∀apply the rules **do**
On the dataset R(d1,d2,d3,..... dn) **end for**
Classify the dataset based on the rules C(R(d_1),R(d_2) .....R(d_n))
Output: Classified datasets with rules C(R(d_1),R(d_2)
.....R(d_n))

Algorithm 3 Apply Classifier on Extracted Features

Apply the hybrid method based on the error rate

$$XF(n) = d + m_1x_1 + m_2x_2 + ... + m_nx_n$$
$$\sum_0^n F(0) = Gain + \sum_0^n w_ix_i$$

C. CLASSIFICATION MODELLING

Dataset clustering is done on the basis of the Decision Tree (DT) features variables and criteria. Instead, to approximate its output, the classifiers are applied to each clustered dataset. Based on their low error rate, the best performing models are identified from the above results. By choosing the DT cluster with a high error rate and extracting its corresponding classifier features, the performance is further optimized. The classifier's output is evaluated on this data set for error optimization.

1) DECISION TREE

The trees are built on the basis of high entropy inputs for training data D samples. In a top-down recursive divide and conquer (DAC) approach, these trees are simple and fast built. Tree pruning is done to remove the samples on D that are irrelevant.

2) LANGUAGE MODEL

The linear form of solution f (x) = mx+b is solved by the following parameters for the given input features xi, yi with input vector xi of data D:

3) SUPPORT VECTOR MACHINE

Support Vector Machine" (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well.

V. EXPECTED RESULT

Models of prediction are created using 13 features and for modeling techniques the accuracy is measured. The best methods of classification can be found here. It compares the accuracy, classification error, precision, F-measure, sensitivity and specificity. The highest accuracy is achieved by HRFLM classification method in comparison with existing methods.

VI. CONCLUSION AND FUTURE ENHANCEMENTS

Identifying the storage of raw cardiac information health data will help to save human lives on a long-term basis and early detection of heart condition irregularities. In this study, machine learning methods were used to process raw data and provide a new and novel heart disease discernment. Heart disease prediction is challenging and very important in the medical field.

REFERENCES

[1] A. S. Abdullah and R. R. Rajalaxmi, ''A data mining model for predicting the coronary heart disease using random forest classifier,'' in *Proc. Int. Conf. Recent Trends Comput. Methods, Commun. Controls*, Apr. 2012 , pp. 22–25.

[2] A. H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori, and M. Abdar, ''Using PSO algorithm for producing best rules in diagnosis of heart disease,'' in *Proc. Int. Conf. Comput. Appl. (ICCA)*, Sep. 2017, pp. 306 – 311.

[3] N. Al-milli, ''Backpropogation neural network for prediction of heart disease,'' *J. Theor. Appl.Inf. Technol.*, vol. 56, no. 1, pp. 131–135, 2013.

[4] C. A. Devi, S. P. Rajamhoana, K. Umamaheswari, R. Kiruba, K. Karunya, and R. Deepika, ''Analysis of neural networks based heart disease prediction system,'' in *Proc. 11th Int. Conf. Hum. Syst. Interact. (HSI)*, Gdansk, Poland, Jul. 2018, pp. 233–239.

[5] P. K. Anooj, ''Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules,'' *J. King Saud Univ.-Comput. Inf. Sci.*,vol.24,no.1,pp.27–40,Jan.2012.doi:10.1016/j.jksuci.2011.09.002.

[6] L. Baccour, ''Amended fused TOPSIS-VIKOR for classification (ATOVIC) applied to some UCI data sets,'' *Expert Syst. Appl.*, vol. 99 , pp. 115–125, Jun. 2018. doi: 10.1016/j.eswa.2018.01.025.

[7] C.-A. Cheng and H.-W. Chiu, ''An artificial neural network model for the evaluation of carotid artery stenting prognosis using a national-wide database,'' in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 2566–2569.

[8] H. A. Esfahani and M. Ghazanfari, ''Cardiovascular disease detection using a new ensemble classifier,'' in *Proc. IEEE 4th Int. Conf. Knowl.Based Eng. Innov. (KBEI)*, Dec. 2017, pp. 1011–1014.

[9] F. Dammak, L. Baccour, and A. M. Alimi, ''The impact ofcriterion weights techniques in TOPSIS method of multi-criteria decision making in crisp and intuitionistic fuzzy domains,'' in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-* R *IEEE)*, vol. 9, Aug. 2015, pp. 1–8.

[10] . Das, I. Turkoglu, and A. Sengur, ''Effective diagnosis of heart disease through neural networks ensembles,'' *Expert Syst. Appl.*, vol. 36, no. 4, pp. 7675–7680, May 2009. doi: 10.1016/j.eswa.2008.09.013.

[11] M. Durairaj and V. Revathi, ''Prediction of heart disease using back propagationMLPalgorithm,''*Int.J.Sci.Technol.Res.*,vol.4,no.8,pp.235–239, 2015.

[12] M. Gandhi and S. N. Singh, ‘‘Predictions in heart disease using techniques ofdatamining,’’in*Proc.Int.Conf.FuturisticTrendsComput.Anal.Knowl. Manage. (ABLAZE)*, Feb. 2015, pp. 520–525.

[13] A. Gavhane, G. Kokkula, I. Pandya, and K. Devadkar, ‘‘Prediction of heart disease using machine learning,’’ in *Proc. 2nd Int. Conf. Electron., Commun. Aerosp. Technol. (ICECA)*, Mar. 2018, pp. 1275–1278.

[14] B. S. S. Rathnayakc and G. U. Ganegoda, ‘‘Heart diseases prediction with data mining and neural network techniques,’’ in *Proc. 3rd Int. Conf. Converg. Technol. (I2CT)*, Apr. 2018, pp. 1–6.

[15] N. K. S. Banu and S. Swamy, ‘‘Prediction of heart disease at early stage using data mining and big data analytics: A survey,’’ in *Proc. Int. Conf. Elect., Electron., Commun., Comput. Optim. Techn. (ICEECCOT)*, Dec. 2016, pp. 256–261.

[16] J. P. Kelwade and S. S. Salankar, ‘‘Radial basis function neural network for prediction of cardiac arrhythmias based on heart rate time series,’’ in *Proc. IEEE 1st Int. Conf. Control, Meas. Instrum. (CMI)*, Jan. 2016, pp. 454–458.

[17] V. Krishnaiah, G. Narsimha, and N. Subhash, ‘‘Heart disease prediction system using data mining techniques and intelligent fuzzy approach: A review,’’ *Int. J. Comput. Appl.*, vol. 136, no. 2, pp. 43–51, 2016.

[18] P. S. Kumar, D. Anand, V. U. Kumar, D. Bhattacharyya, and T.-H. Kim, ‘‘A computational intelligence method for effective diagnosis of heart