HEART DISEASE PREDICTION: A REVIEW

Treesa Maria Pius[1] ,Akhila Madhu[2], Arun Antony[3], Lisha Varghese[4]

[1,2,3]P.G Scholar of Amal Jyothi College of Engineeing, Kanjirappally, Kerala

[4]Asst. Professor of Amal Jyothi College of Engineering, Kanjirappally, Kerala

Abstract- Investigation of relevance of medical and demographic information helps in predicting the presence of heart disease in an individual. Applying a variety of ensemble and deep learning techniques with hyper parameter tuning and feature selection, resulting in a maximum test accuracy of 78%. Model averaging does not significantly improve prediction accuracy, and the same points tend to be misclassified by all the models that don't overfit. This suggests that the errors in our data can mainly be attributed to an irreducible error in the problem [5]. Cardiovascular diseases are the most common cause of death worldwide over the last few decades in the developed as well as underdeveloped and developing countries. Early detection of cardiac diseases and continuous supervision of clinicians can reduce the mortality rate. However, accurate detection of heart diseases in all cases and consultation of a patient for 24 hours by a doctor is not available since it requires more sapience, time and expertise [1].

Keywords- Data Mining, Machine Learning, IoT (Internet of Things), Artificial Intelligence (AI), Particle Swarm Optimization (PSO).

I. INTRODUCTION

The heart is important organ or part of our body. Life is itself dependent on efficient working of heart. If operation of heart is not proper, it will affect the otherbody parts ofhuman such as brain, kidney etc. It is nothing more than a pump, which pumps blood through the body. If circulation of blood in body is inefficient the organs like brain suffer and if heart stops working altogether, death occurs within minutes [7].

Cardiovascular disease generally refers to conditions that involve narrowed or blocked blood vessels that can lead to a heart attack, chest pain (angina) or stroke. Other heart conditions, such as those that affect your heart's muscle, valves or rhythm, also are considered forms of heart disease. Cardiovascular disease symptoms may be different for men and women. For instance, men are more likely to have chest pain; women are more likely to have other symptoms along with chest discomforts, such as shortness of breath, nausea and extreme fatigue.

Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves. The iterative aspect of machine learning is important because as models are exposed to new data, they are able to independently adapt. In statistics and machine learning, ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone Unlike a statistical ensemble in statistical mechanics, which is usually infinite, a machine learning ensemble consists of only a concrete finite set of alternative models, but typically allows for much more flexible structure to exist among those alternatives.

Prediction refers to the output of an algorithm after it has been trained on a historical dataset and applied to new data when you're trying to forecast the likelihood of a particular outcome, such as whether or not a customer will churn in 30 days. The algorithm will generate probable values for an unknown variable for each record in the new data, allowing the model builder to identify what that value will most likely be [6].

II. LITERATURE REVIEW

Nashif et al [1], data mining provides a number of techniques which discover hidden patterns or similarities from data. Therefore, in this paper, a machine learning algorithm is proposed for the implementation of a heart disease prediction system which was validated on two open access heart disease prediction datasets. Another contribution of this paper is the presentation of a cardiac patient monitoring system using the concept of the IoT with different physiological signal sensors and Arduino microcontroller. Sensor networks are currently using the IoT technology to collect, analyze and passing of information from one node to another. The sensors collect the data after a specific time, analyze it and use it to initiate the required action, and provide an intelligent cloud-based network for analysis, planning and decision making.

Miaoa et al [2], an advanced deep neural network approach is developed and utilized to predict coronary heart disease in patients and increase diagnostic accuracy using classification and prediction models based on deep learning. For this research, the developed classification and diagnosis models contain two parts: a deep neural network learning-based training model and a prediction model for the presence of heart disease. The training model is first created using deep learning algorithms based on a deeper multilayer perceptron with regularization and dropout in system and architecture. Based on the training model, the diagnosis model is then utilized to predict whether or not patients have coronary heart disease. The subsequent performance of the deep learning model for heart disease diagnosis is evaluated in terms of the performance measure parameters, including diagnostic accuracy, probability of misclassification error, sensitivity, specificity, precision, area under the ROC curve (AUC), Kolmogorov-Smirnov (K-S) measure, receiver operating characteristic (ROC), and F-score.

Khourdifi et al [3], the PSO algorithm has been successfully applied to heart disease because of its simplicity and generality. However, PSO easily fell into the optimal local solution. ACO algorithms have been developed to solve continuous optimization problems. These problems are characterized by the fact that decision variables have continuous domains, unlike discrete problems. To further explore the application of intelligent optimization in bioinformatics, PSO and ACO are combined in this article, meaning that exploitation and exploration capacity are combined for binary and multi-class heart disease. In this article, the Fast Correlation-Based Feature selection (FCBF) method used to remove redundant and irrelevant features, the results of the PSO optimization are considered the initial values of the ACO, and then the classification model for heart disease is constructed after the parameters are adjusted.

Ul Haq et al [4], the contribution of the proposed research is to design a machine-learning-based medical intelligent decision support system for the diagnosis of heart disease. In the present study, various machines learning predictive models such as logistic regression, k-nearest neighbor, ANN, SVM, decision tree, Naive Bayes, and random forest have been used for classification of people with heart disease and healthy people. Three feature selection algorithms, Relief, minimal redundancy-maximal-relevance (mRMR), Shrinkage and Selection Operator (LASSO), were also used to select the most important and highly correlated features that great influence on target predicted value. Cross-validation methods like k-fold were also used. In order to evaluate the performance of classifier, various performance evaluation metrics such as classification accuracy, classification error, specificity, sensitivity, Matthews' correlation coefficient (MCC), and receiver optimistic curves (ROC) were used.

III. IMPLEMENTATION

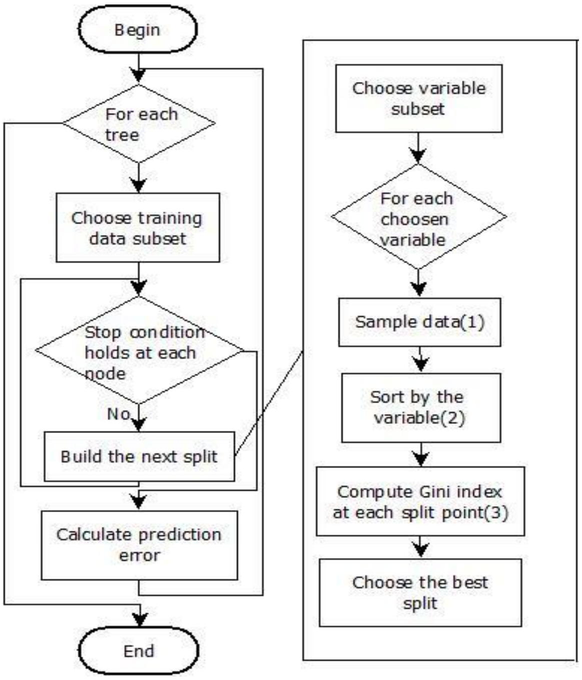
A compiled dataset of 899 anonymous patients from three hospitals: Cleveland Clinic, University Hospital of Switzerland, and the Hungarian Institute of Technology. These patients were all suspected of having heart disease, and a little less than 60% of them actually did. These patients were each tested for heart disease via an electrocardiogram and an exercise test, as well as more advanced procedures like X-ray fluoroscopy and blood tests.

Then removed X-ray fluoroscopy and blood test information from the data, leaving only features that can be measured without the assistance of a medical professional. The further augmented data with aggregate metrics from a government study that detailed heart disease frequencies between different age buckets and the two sexes [5].

Random forest is well-known for taking care of data imbalances in different classes especially for large datasets. Due to its parallel architecture, random forest classifier is faster compared to other state-of-the-art classifiers [8].

Random Forest works as follows:

1. Selects k features (columns) from the dataset (table) with a total of m features randomly (where k<<m). Then, it builds a Decision Tree from those k features.
2. Repeats n times so that you haven Decision Trees built from different random combinations of k features (or a different random sample of the data, called bootstrap sample).
3. Takes each of the n built Decision Trees and passes a random variable to predict the outcome. Stores the predicted outcome (target), so that you have a total of n outcomes from then Decision Trees.
4. Calculates the votes for each predicted target and takes the mode (most frequent target variable). In other words, considers the high voted predicted target as the final prediction from the random forest algorithm.



IV. METHOD OF IMPLEMENTATION

An interface of this research element has been developed. The interface of this element is been made using a MySQL database for the entering values of patients. The backend of the element is developed on the Python Programming Language using flask framework.

Random Forest Classifier –Prediction

```
scaler_path=
os.path.join(os.path.dirname(__file__),
'models/scaler.pkl')
scaler = None
with open(scaler_path, 'rb') as f:
scaler = pickle.load(f)
x = scaler.transform(x)
model_path=
os.path.join(os.path.dirname(__file__),
'models/rfc.sav')
clf = joblib.load(model_path)
y = clf.predict(x)
print(y)
```

V. FUTURE WORKS

Based on findings, the hypothesis that increased training data will help to reduce the misclassification rate. So here plan on collecting additional, high-quality patient data from diverse resources and also hope to extend work to include not only diagnosing heart disease, but also predicting its likelihood of on set in a given period of time in the future. The plan on testing the most predictive features discovered in the current project to see how well it can be fare in predicting future heart disease.

VI. CONCLUSION

It has been suggested that the peak prediction accuracy of significant work still needs to be done in order to reach a performance comparable to that of a medical professional. However, it has been believed that the initial findings demonstrate the possibility of reaching that level of accuracy, especially by making use of features determined to be most predictive of heart disease within the dataset.

REFERENCES

[1] ShadmanNashif , "Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System", World Journal of Engineering and Technology, 2018, 6, 854-873.

[2] Kathleen H. Miaoa , "Coronary Heart Disease Diagnosis using Deep Neural Networks", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9, No. 10, 2018.

[3] YounessKhourdifi, "Heart Disease Prediction and Classification Using Machine Learning Algorithms Optimized by Particle Swarm Optimization and Ant Colony Optimization", International Journal of Intelligent Engineering &Systems, October 22, 2018.

[4] Amin Ul Haq, "A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms", Mobile Information Systems V, Article ID 3860146, 21 pages Volume 2018.

[5] BoyangDun , "Heart Disease Diagnosis on Medical Data Using Ensemble Learning", Vol. 1, No. 1, Article 1. Publication date: January 2016.

[6] Elham Nikookar, "Hybrid Ensemble Framework for Heart Disease Detection and Prediction", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9, No. 5, 2018.

[7] K.Sudhakar, "Study of Heart Disease Prediction Using Data Mining", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 1, January 2014.

[8] Angshuman Paul, "Improved Random Forest for Classification",IEEE Transactions on Image Processing (Vo lume: 27 , Issue: 8 , Aug. 2018)