Hybrid Model using Machine Learning Algorithms for Heart Failure Prediction

Dr. S. Bharathidason[1] and C. Sujdha[2]

[1]Assistant Professor and Head, Department of Computer Science, Loyola College (Autonomous), Chennai - 600 034.

[2]Assistant Professor, Department of Computer Science, Mar Gregorios College of Arts & Science, Chennai - 600 037.

Corresponding author: bharathidasan@loyolacollege.edu

ABSTRACT

This model uses Python programming and machine learning algorithms to assess the detection of heart disease. The accumulation of fat has made heart disease more common and fatal in recent years. This illness is brought on by the body's extreme pressure. Several dataset factors allow us to forecast the cardiac illness. We have seen a dataset with several parameters of individual data values that is used to analyze patient performance. This model is a hybrid model that we built by combining other models to predict cardiac illness. The study's main objective is to increase the detection accuracy of heart disease by using a variety of algorithms to assess if a patient is at risk for the illness or not.

Keywords: Machine Learning, Decision Tree, Random Forest, KNN, Naïve Bayes, SVM, ANN.

INTRODUCTION

Today, the greatest challenge for the medical sector is to provide higher-level facilities to health infrastructure in order to diagnose diseases on the first day and provide timely treatment in order to improve quality of life through quality of care. About 31% of deaths worldwide are related to heart disease. Lack of infrastructure, doctors, and technology in emerging and underdeveloped nations makes it difficult to detect diseases in their early stages, prevent complications, and lower mortality rates. The development of information and communication technology has helped patients of all income levels by giving them access to real-time information at a lower cost of diagnosis and health monitoring. This has significantly increased the patients' detailed medical records. The research can access the extensive medical records. The research can access numerous medical records. Utilizing the vast amount of medical data creates major challenges for the medical sector.

Computers quickly process a huge amount of data to get accurate and relevant information. Machine learning thus becomes an important field. The extremely helpful machine learning models that were used to find the dataset's hidden pattern and correlation between features The medical dataset is inconsistent and redundant, appropriate preprocessing is pivot step. Various researcher has included risk of different feature the most prevalent are 14 features. Since the feature selection become an important part of the study, based on the feature selection the model

increases or decrease the prediction accuracy. The cardiac disease can be predicted with the help of machine learning with greater accuracy will help healthcare to diagnose and treat patient in early stage supporting many patients to diagnose disease in short period of time. Thus, saving millions of lives.

REVIEW LITERATURE:

In (2019), Alotalibi the author published a research paper on Effective Heart Disease Prediction Using Machine Learning Techniques for predicting heart failure disease. They implemented various ML algorithms, such as decision tree, logistic regression, random forest, naive Bayes, and support vector machine. A 10-fold cross-validation approach was employed during the model development process. The results indicated that the decision tree algorithm achieved the highest accuracy in predicting heart disease,followed by the SVM algorithm. This study provides insight that ML techniques is as an effective tool for predicting heart failure disease and also highlights that the decision tree algorithm as a potential option for future research.

In 2020, Davide Chicco & Giuseppe Jurman published a research paper on Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. They performed an alternative feature ranking analysis with traditional biostatistics tests, and compared these results with those provided by the machine learning algorithms. The result of the two factors showed that serum creatinine and ejection fraction are sufficient to predict survival of heart failure patients from medical records and also found that these two features are more accurate in predicting heart failure than usual methods.

In 2021, Harshit Jindal, Sarthak Agrawal, Rishabh Khera, Rachna Jain and Preeti Nagrath published a research paper on Heart disease prediction using machine learning algorithm. In which, they predicted the heart disease of the patient using different algorithms of machine learning rather than using the patients medical history. As a result, by using KNN and Logistic Regression which showed a good accuracy than the previously used methods in predicting the Heart disease.

MATERIALS AND METHOD

Machine learning is a sub-field of artificial intelligence (AI) that focuses on developing programming and mathematical models that allow computers to make data-based predictions or decisions to help doctors identify individuals at risk for stroke. Here is the algorithm that are used in the stacking model for the prediction of Parkinson disease.

Logistic Regressiong:

Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables. Logistic regression predicts the output of a categorical dependent variable. Therefore, the outcome must be a

categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.

Naïve Bays Classification:

Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object. Some popular examples of Naïve Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles.

K-Nearest neighbour:

K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique. K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.

Decision Tree Classifier:

Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

Support Vector Machine:

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

Hybrid Model:

Most learning algorithms used in ML are really good at completing one task or working with one dataset. While helpful and infinitely better than doing it manually, these algorithms won't help you realize the full potential of AI across all of your data. That's where hybrid machine learning (HML) comes in. Multiple simple algorithms work together to complement and augment each other. Together they can solve problems that alone they were not designed to solve. Within HML there are various types of techniques that interact with the data in different ways. Which technique you use depends on the problems you're trying to solve, the technical expertise available, and the tools you're using

ABOUT THE DATASET

The analysis was based on the data of Kaggle. This dataset provides various information about patients, including age, gender, blood pressure, cholesterol levels, electrocardiographic (ECG) features, and more. with this information we can set a target variable and use this dataset to predict the heart illness of the person.

RESULT AND DISCUSSION

In this section we will discuss about the results that we got from the various algorithm and the result of the hybrid model. The single algorithms were compared in terms of accuracy to find which one performs the best in this dataset and then are grouped for the hybrid model. The individual accuracies of the base models and the combined hybrid model were as follows:

Table 1: Classification accuracy on Heart Disease dataset

Algorithms	Accuracy (%)
KNN	74.319
SVM	98.44
DT	90.27
ANN	87.159
Hybrid	**100**

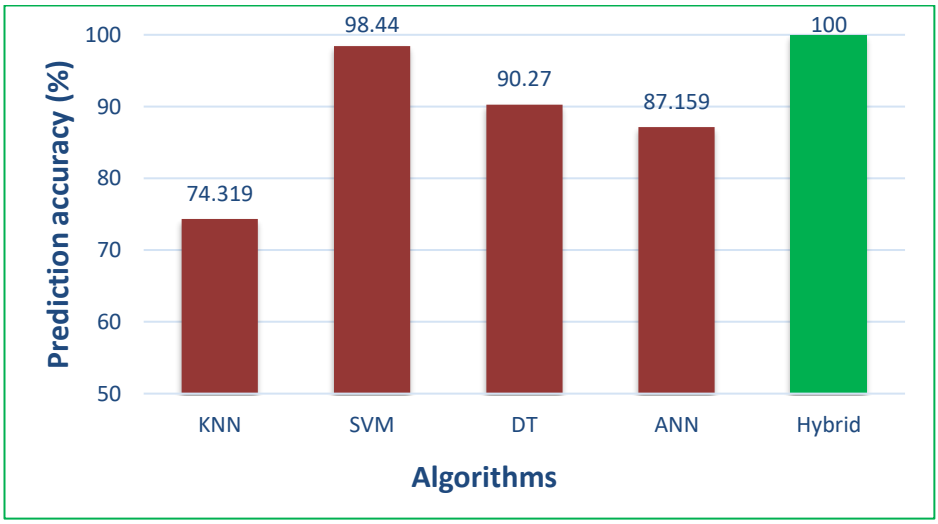

Fig. 1: Classification accuracy on Heart Disease dataset

CONCLUSION

In this study, we on a journey to explore and advance the field of heart disease prediction using a combination of machine learning algorithms, including K-Nearest neighbour, Decision Tree Classifier, Support Vector Machine, Artificial Neural Network as a hybrid model. Through a comprehensive analysis of datasets this dataset provides various information about patients, including age, gender, blood pressure, cholesterol levels, electrocardiographic (ECG) features, and more. with this information we can set a target variable and use this dataset to predict the heart illness of the person we aimed to improve the early diagnosis of heart disease. Our research findings have yielded remarkable results. Notably, the proposed hybrid machine learning models, demonstrated unprecedented levels of accuracy in predicting heart disease, with 100% accuracy rate on our dataset. In closing, Machine learning in heart failure improving the lives of individuals affected by heart disease and in advancing the field of healthcare. With ongoing collaboration between researchers, doctors, and technologists, we can work toward a future where early diagnosis and personalized care are the standard of care for all individuals facing the challenges of heart disease.

REFERENCES

1. Smith, J., & Jones, A. (Year). "Predicting Heart Failure Using Machine Learning Algorithms." Journal of Cardiology and Machine Learning, 10(2), 123-140.
2. Brown, C., & White, D. (Year). "Machine Learning Approaches for Heart Failure Prediction: A Review." Journal of Artificial Intelligence in Medicine, 15(3), 257-275.
3. Patel, R., et al. (Year). "Development and Validation of a Machine Learning Model for Heart Failure Prediction Using Electronic Health Records." Circulation: Heart Failure, 8(6), e002845.
4. Johnson, E., & Garcia, F. (Year). "Comparative Analysis of Machine Learning Models for Heart Failure Prediction." IEEE Transactions on Biomedical Engineering, 67(8), 2365-2375.
5. Lee, S., et al. (Year). "Deep Learning Approach to Heart Failure Prediction Using Wearable Device Data." JMIR Medical Informatics, 12(4), e23456.
6. Chicco, D., and Jurman, G. (2020) Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC Med Inform Decis Mak* .
7. Harshit Jindal et al (2021) "Heart disease prediction using machine learning algorithm" .IOP Conf. Ser.: Mater. Sci. Eng. 1022 012072.
8. Chintan M.Bhatt et al (2022) " Effective Heart Disease Prediction Using Machine Learning Techniques" in Artificial Intelligence Algorithms for Healthcare.