A Novel Approach for Heart Disease prediction using Artificial Intelligence Techniques

V.Sathyavathy [1]

Sathyavathy.v@kgcas.com[1]

Assistant Professor, Department of Computer Technology, KG College of Arts and Science[1]

Abstract. Despite major improvements in diagnosis and care, cardiovascular disease remains to be the world's major cause of morbidity and mortality. Due to the growth in bulk amount of data and complexity, Artificial Intelligence techniques like machine learning approaches and deep learning algorithms can help advance medical understanding by revealing data that is clinically useful. Many medical conditions can be identified, detected, and predicted using machine learning.This main goal of the study is to provide the best algorithmic implementation for early heart disease issue detection. As a result, it will be simpler to provide patients with the right care while minimising negative consequences. In this kind of condition, the heart often has trouble pumping enough blood to the rest of the body so that it can carry out its intended activities. The ability to diagnose this issue quickly and accurately is critical for losing patient's lives and reducing further damage. On the other hand, antagonistic based approaches are suggested to be more accurate and dependable for the detection of heart illness, such as computational algorithms based on intellectual learning. For identifying and diagnosing heart illness, an intelligent computational predictive system is presented in this article.

Keywords: Cardiovascular Disease, Random Forest Algorithm, Artificial Intelligence, Logistic Regression.

1 Introduction

Cardiac failure is the leading cause of death worldwide and is becoming an increasingly difficult problem. The number of Heart Disease cases worldwide has progressed from 271 million in 1990 to 523 million in the 2019 and over the similar period, death has attained more from 12.1 million to 18.6 million. To lessen the burden of CVD, scalable and affordable solutions are required, and Artificial Intelligence (AI) might be a key component [1].

2 AI in Health Care

Large volume of data sets may now be quickly and perfectly analysed to make enhancements in speed and processing capability. This has made it possible for medical practitioners to apply

IACIDS 2023, November 23-25, Lavasa, India

DOI 10.4108/eai.23-11-2023.2343179

AI to vast, complex data sets to improve decision-making, diagnosis and support by looking for patterns in patient data and assisting those who have had a stroke.

2.1 Preventing heart problems

Using AI to examine ECGs has yielded a low-cost test that is generally applicable for identifying the existence of a weak heart pumping blood to the vessels, which, if untreated, can result in cardiac failure. The clinic of Mayo is in a fantastic position to extend this use of AI as it has a database of more than 7 million ECGs. All individually specific patient information is first deleted to preserve privacy.

2.2 Detecting fibrillation a trialsooner

Artificial Intelligence-directed ECGs are also utilised to detect abnormal cardiac pulses (atrial fibrillation) before to the presence of any symptoms.

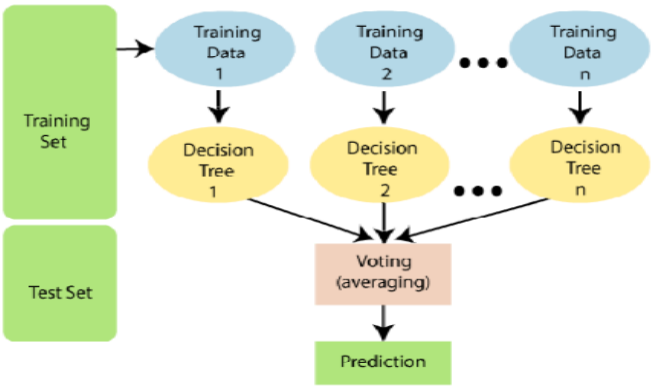

Fig.1. Training Set and Testset

3 Literature Review

Artificial Intelligence allows the human to interact and able to solve complex problem through smart machines.

There are various models that evaluates the heart disease risk prediction such as logistic regression, KNN, and random forest. The evaluation result of each algorithm performs a good accuracy of 87.6%

This paper performs a comprehensive survey of various models that analyses the algorithms and performance. The study uses RF, LR, KNN classifiers to provide the base for the beginners. Layer based learning rate is incorporated in the function of the meta learning. Finally, the recommended stacking model is evaluated and considered to its precision, accuracy, recall, F1 score. In order to assess the flexibility of the model to new circumstances, a dataset which is available known as the Heart Disease Dataset [5].

Table 1. Dataset

Attribute	Meaning
Age	Age is Continuous
Gender	1=male 0=female
Cp	Chest Pain
Trestbps	Resting Blood Pressure(mmHg)
Chol	Cholesterol
Fbsl	Fasting blood sugar 0:<=120mg/dl, 1:>120mg/dl
restecg	electrocardiographic results during resting 1=true 0=false
thalach	Maximum heart rate achieved: continuous
exang	Exercise induced angina
oldpeak	ST depression
slope	ST segment slope
ca	Niumber of major vessels coloured by fluoroscopy:discrete(0,1,2,3)
thal	3: normal 6: fixed defect 7: reversible defect

4 Random Forest Algorithm

This algorithm is the most potential and popular in machine learning. It is a component of supervised machine learning. This algorithm is implemented for both regression and classification issues. This technique gathers data, creates decision-based trees based on several datas, then the average is taken for the decision trees. Problems involving regression and classification may both be solved using this technique.

5 Logistic Regression (LR)

For studying datasets with links between the independent factors that impact a result and a categorised dependent variable (DV), logistic regression is one of the machine learning classification methods. Using a variety of value identifiers, the Logistic Regression (LR) Classifier forecasts the outcome of a dependent variable. It is suitable when the autonomous variables are all out, persistent, or merged, and the dependent component is all out and dichotomous. Using logistic regression, there are two ways of demonstration:

1. Regression stepwise - A regression approach where the significance of the model is evaluated after each autonomous component is inserted.

2. Regression in Reverse stepwise – A strategy that uses all available independent components and gradually eliminates unrelated features in order to check that the model adequately matches the data.

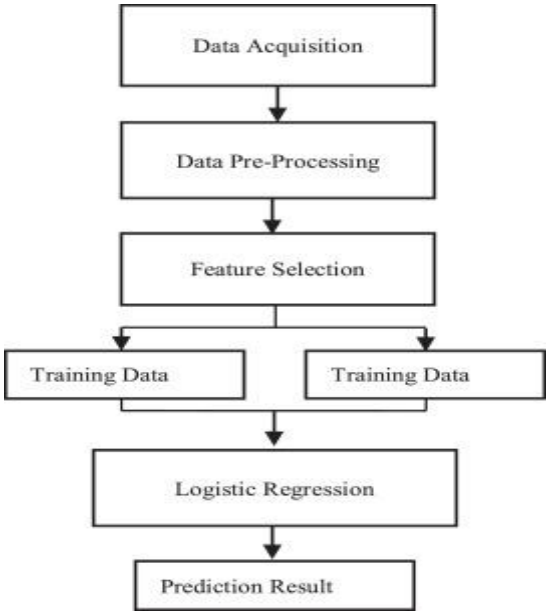

Fig.2. Logistic Regression Approach

6 KNN Algorithm

In Artificial Intelligence techniques, the KNN algorithm is simplest and supervised learning algorithms. The K-Nearest Neighbor algorithm classifies the parameters in the dataset of test by differentiating with the observations in the dataset of training the performance of the KNN model is assessed.

$$Average\ Accuracy = \sum_{i=1}^{l} \frac{TP_i + TN_i}{TP_i + FN_i + FP_i + TN_i} / l$$

Fig.3. KNN Formula

7 Machine Learning Model's performance Analysis

Following the preparation of the heart disease dataset, a number of classification methods were used utilising 10-fold cross-validation approaches, including KNN, Logistic Regression, Random Forest, Different cross-validation measures were used to evaluate each algorithm's performance, and the top algorithm was determined. The following table illustrates the complete procedure and lists the classifiers evaluation metrics. AUROC, accuracy, F1, precision, and recall were chosen as the evaluation measures to give a detailed analysis of each performance of the classifier and to provide an equitable comparison of their performance.

1.Random Forest Classifier

The analysis of this classification training accuracy is 88% and the testing accuracy is 85%.

Table 2.Accuracy & Performance – Random Forest

Accuracy for Testing				
Heart Disease	Precision	recall	F1-score	Support
0	0.92	0.86	0.88	35
1	0.73	0.88	0.89	19
Performance				
Macro avg	0.81	0.88	0.86	56
Weighted avg	0.89	0.86	0.88	54
Accuracy			0.85	54

Accuracy for Training				
Heart Disease	Precision	recall	F1-score	Support
0	0.88	0.90	0.87	110
1	0.89	0.85	0.89	100
Performance				
Macro avg	0.88	0.88	0.88	216
Weighted avg	0.88	0.88	0.88	216
Accuracy			0.88	216

2. Logistic Regression

For the Logistic Regression in the model training accuracy is 85% and the testing accuracy is 81%

Table 3. Performance & Accuracy – Logistic Regression

Accuracy for Training				
Heart Disease	Precision	recall	F1-score	Support
0	0.81	0.87	0.84	116
1	0.89	0.83	0.82	100
Performance				
Macro avg	0.85	0.85	0.85	0.85
Weighted avg	0.85	0.85	0.85	216
Accuracy			0.85	216

Accuracy for Testing				
Heart Disease	Precision	recall	F1-score	Support
0	0.96	0.75	0.84	37
1	0.63	0.93	0.74	16
Performance				
Macro avg	0.80	0.85	0.81	55
Weighted avg	0.87	0.82	0.85	54
Accuracy			0.85	54

2. KNN Algorithm

The accuracy is 74% in the training model of KNN Algorithm and 74% in the testing phase.

Accuracy for Testing				
Heart Disease	Precision	recall	F1-score	Support
0	0.83	0.75	0.81	38
1	0.55	0.70	0.62	16
Performance				
Macro avg	0.72	0.73	0.73	54
Weighted avg	0.75	0.74	0.75	54
Accuracy			0.74	54

Accuracy for Training				
Heart Disease	Precision	recall	F1-score	Support
0	0.74	0.77	0.75	110
1	0.74	0.76	0.72	106
Performance				
Macro avg	0.74	0.73	0.74	216
Weighted avg	0.74	0.74	0.74	216
Accuracy			0.74	216

Table 4. Accuracy & Performance – KNN

The findings of the study demonstrate that ANN offers maximum. The study's findings demonstrate that Random Forest outperforms Logistic Regression, KNN, Random Forest. In order to help physicians estimate the likelihood that their patients may develop heart disease, our study discovered many ML classifiers with the capacity to do so. To create a more reliable prediction model, further data and analysis are needed as the dataset utilised in this study included little information on heart disease. Despite this drawback, it is believed that future studies will enhance our comprehension of the advantages and restrictions of this strategy and that applying machine learning algorithms to the examination of new data.

Finally ROC Curve is created to show the performance of model's at various levels of thresholds.

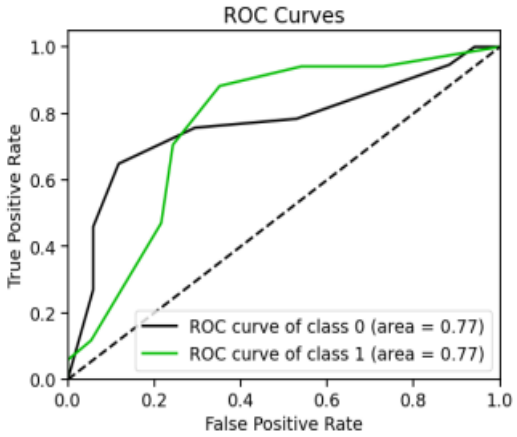

Fig.4. ROC Curve

7 Conclusions

Heart disease is a serious ailment that can result in fatal consequences, such as cardiac attacks, as well as life-threatening complications. This study sought to assess the usefulness of ML methods in cardiac disease prediction using a dataset of heart disease. The techniques of Machine Learning have the ability to correctly predict the onset of heart disease. To ensure impartial methods of machine learning, the data was pre-processed to deal with lost values, normalised data, and compensate for imbalance. In conclusion, using a variety of feature selection strategies, the most important features that are extremely beneficial for predicting heart disease was identified, and then applied six different machine learning algorithms for those features [10]. Each algorithm generated a unique score based on a different set of criteria. The

performance of Random forest was more noteworthy than that of the other algorithms. More research will be done in the future to identify more effective algorithms, such as deep learning algorithms, to realise this promise of increased performance of the algorithms utilising more efficient feature selection strategies.

References

[1] World Health Statistics. Cardiovascular Diseases, Key Facts. 2021. Available online: https://www.who.int/news-room/factsheets/detail/cardiovascular-diseases-(cvds) (accessed on 10 December 2022).

[2] Choudhury, R.P.; Akbar, N. Beyond Diabetes: A Relationship between Cardiovascular Outcomes and Glycaemic Index. Cardiovasc.Res. 2021, 117, E97–E98.

[3] Magesh, G.; Swarnalatha, P. Optimal Feature Selection through a Cluster-Based DT Learning (CDTL) in Heart Disease Prediction.Evol. Intell. 2021, 14, 583–593.

[5] Rohit Chowdary, K.; Bhargav, P.; Nikhil, N.; Varun, K.; Jayanthi, D. Early Heart Disease Prediction Using Ensemble LearningTechniques. J. Phys. Conf. Ser. 2022, 2325, 012051.

[6] Liu, J.; Dong, X.; Zhao, H.; Tian, Y. Predictive Classifier for Cardiovascular Disease Based on Stacking Model Fusion. Processes2022, 10, 749.

[7] Devi, A.G. A Method of Cardiovascular Disease Prediction Using Machine Learning. Int. J. Eng. Res. Technol. 2021, 9, 243–246.

[8] Uddin, S.; Khan, A.; Hossain, M.E.; Moni, M.A. Comparing Different Supervised Machine Learning Algorithms for DiseasePrediction. BMC Med. Inform. Decis. Mak. 2019, 19, 281.

[9] Patro, S.P.; Nayak, G.S.; Padhy, N. Heart Disease Prediction by Using Novel Optimization Algorithm: A Supervised LearningProspective. Inform. Med. Unlocked 2021, 26, 100696.

[10] Pasha, S.J.; Mohamed, E.S. Novel Feature Reduction (NFR) Model with Machine Learning and Data Mining Algorithms forEffective Disease Risk Prediction. IEEE Access 2020, 8, 184087–184108.