ISSN: 2633-4828 Vol. 5 No.4, December, 2023

International Journal of Applied Engineering & Technology

GUI BASED HEART DISEASE PREDICTION MODEL USING RANDOM FOREST ALGORITHM

Mr. S. S. Mhaske[1] and Dr C. M. Jadhao[2]

[1]Research Scholar and [2]Principal & Professor, Mauli Group of Institutions, College of Engineering and Technology, Shegaon, Maharashtra, India

ABSTRACT

Cardiac disease is the primary cause of death globally. Diagnosis and accurate prediction of cardiac illness are difficult tasks. The mortality rate due to cardiac disease can be significantly decreased by routine monitoring and early detection of cardiac disease. The medical industry is seeing a considerable advancement in the application of machine learning techniques. Improved accuracy in predicting cardiac diseases has been obtained by interpreting the analyses done using this methodology. It has been demonstrated to explicitly attempt to uncover significant aspects of heart disease by giving ML. The classification algorithm Random Forest (RF) is used to train the model.

Keywords: Cardiac disease, Machine learning, Random Forest.

INTRODUCTION

An essential organ in the human body is the heart. It supplies blood to every bodily part. In the event that it malfunctions, the brain and several other organs will stop functioning, and the person will pass away in a matter of minutes. Any ailment that affects the heart might be collectively referred to as heart disease. Heart disease encompasses several disorders such as heart valve disease, cardiomyopathy, coronary heart disease, elevated blood pressure, cardiovascular disease, heart failure on one side, pulmonary failure, and more conditions. Heart disease risk factors can be acquired or inherited, and they can manifest at any age. Additionally, eating foods high in fat, like eggs or trans fats, smoking, overindulging in junk food, being overweight, abusing alcohol or soft drinks, living a sedentary lifestyle, diabetes, smoking, high blood pressure, and obesity are all contributing factors.

As per the article published by The Economic Times on September 29, 2023 [13], in India, cardiovascular diseases have turned into a silent epidemic of sorts and account for at least 27 percent of deaths.

LITERATURE REVIEW

The paper published by Sree et al., 2023 [1], presented a comprehensive analysis of risk prediction for heart disease using four machine learning algorithms. They had trained the model using four classification algorithms: Decision Tree (DT), K-Nearest Neighbors (K-NN), Random Forest (RF), and Support Vector Machine (SVM). They had measured four algorithms performance in a number of ways, including recall, accuracy, precision, and specificity. It has been demonstrated that SVM offers the best performance in this approach.

Jummelal, 2023 [2], presented a machine-learning method for identifying chronic heart failure using ECG recordings. The method, which includes filtering, segmenting, and feature extraction, outperforms a majority classifier by 15 percentage points and has an 87% recall rate among those with chronic heart failure.

The work presented by Alkurdi, 2023[3], used the Heart Disease UCI dataset to train machine learning models for heart disease diagnosis. Using preprocessing techniques like MEAN, normalization, SMOTE, and correlation, the data is pre-processed, and then this pre-processed data is forwarded to decision Tree, Random Forest, SVM, and k-NN classification algorithms. The methodology that has been suggested yields encouraging outcomes that distinctly underscore the importance and worth of preparing data. The obtained accuracy, precision, recall, F1 score, and ROC AUC scores all clearly demonstrate this.

Swain et al., 2023 [4], employed IoT to assess risk variables that are comparable to improperly managed cardiac conditions. They said that there has been a significant technological divide between researchers studying

healthcare, patients, doctors, and machine learning architects. They have conducted an intense review of machine learning with the aim of taking advantage of new, cutting-edge technologies.

The research work presented by Swain et al., 2023[5], describes how an ML model can predict whether or not an individual would be at risk of CHD in the following ten years based on medical data and personal habits. They aim to determine the most accurate model for heart disease prediction using a variety of machine learning (ML) classification techniques, including random forests (RF), logistic regression (LR), artificial neural networks (ANNs), and decision trees (DT), where they have found out that RF has the best accuracy.

Zabeeulla et al., 2023[6], presented a machine learning algorithm to diagnose cardiac problems in which they achieved 98.8% accuracy. They have used a heart disease data set from Kaggle, which is open to all, and the dataset contained thirteen features. Finding out if the patient has a 10-year risk of developing coronary heart disease" (CHD) was the ultimate goal of their study.

To ascertain the effectiveness of various machine learning algorithms in categorization tasks, a thorough investigation is carried out by Hadianti, 2023 [7]. The experimentation dataset comes from a variety of patient populations, which increases the findings' generalizability. The study aims to determine the best machine learning approach for accurately predicting cardiac disease through thorough review and validation. Using Optuna's optimization technique, the author was able to increase the accuracy of heart disease forecasts.

Nayeem et al., 2022[8], used supervised machine learning algorithms to predict heart disease in patients. They had employed mean value and info-gain feature selection techniques to handle null values and remove unnecessary features. K-Nearest Neighbors, Naive Bayes, and Random Forest are used to calculate prediction accuracy, precision, recall, F1-score, and ROC. Out of all the models, Random Forest has the highest classification accuracy (95.63%), with the following values for precision, recall, F1-score, and ROC: 0.93, 0.92, 0.92, and 0.9.

The authors, Rindhe et al., 2021[9], performed the simulation on a dataset available in the UCI Machine Learning repository, which consists of 303 samples with fourteen input features and one output feature. They have used supervised machine learning algorithms like random forest and support vector machine as well as Artificial Neural Network. The web scraper responsible for gathering data for the model is programmed in Python. The accuracy achieved by them is 84.0%, 83.5%, and 80.0% for Support Vector Classifier, Neural Network, and Random Forest Classifier respectively.

Ruqiya, 2023[10], presented detailed review of the Cleveland Heart Disease Dataset using Machine Learning, and from the review, they suggested increasing the number of samples in the dataset for evaluation and implementing the deep learning approach after increasing the number of samples in the data set, as well as considering more features for classification.

The author, Handan Tanyildizi-Kökkülünk, 2023 [11], presented the work on Prediction of Heart Disease Using Machine Learning with Data Mining. Using multiple linear regressions (MLR), a machine learning technique, the categorization was achieved using the R Studio software. By employing the Akaike information criterion to identify characteristics that significantly contribute to the prediction, machine learning has been enhanced. The MLR model that was employed had an accuracy of 88%, a precision of 93%, a sensitivity of 86%, and a specificity of 91%.

Anusuya & Gomathi, 2021[12], presented a method that involved preprocessing, followed by feature selection, Eigen vector extraction, and classification. Moreover, multi-objective-based Ant Colony Optimization (MO-ACO) is used to choose the best features from the extracted features in order to increase the classification and clustering. The suggested approach was contrasted with the current methodology in terms of accuracy, recall, precision, NMI, and execution time, and came to the conclusion that the solution was more accurate for datasets of all sizes.

METHODOLOGY:

The system flow for heart disease prediction is as shown in Figure 1. It involves steps like data preprocessing of the input data set, feature extraction, feature selection, splitting of data into test and train data sets, classification, and the output of the classifier, which is the prediction of heart abnormality.

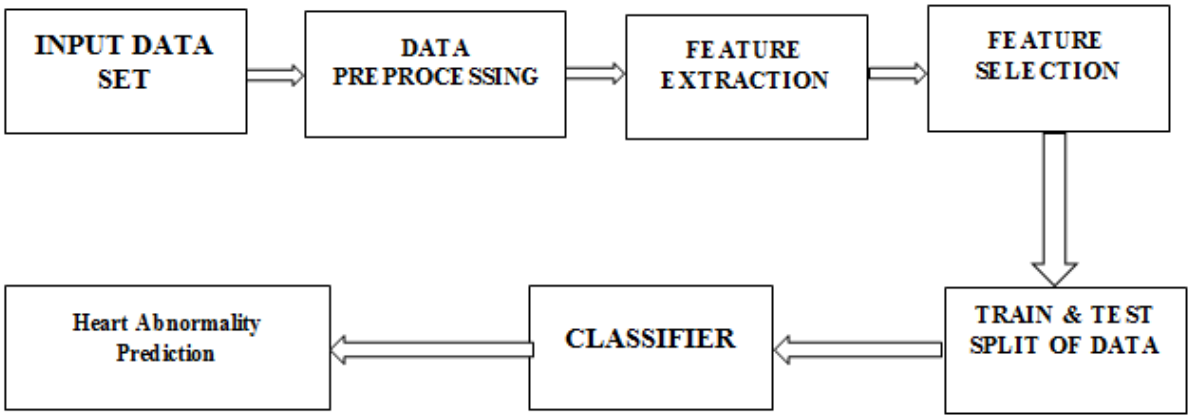

Fig 1: System flow for prediction of heart disease

The online data set from the github website [14] is used, which has 1025 data samples, and these data samples are used to train the random forest classifier. The various parameters like age, sex, CP, trestbps (the person's resting blood pressure), cholesterol, fbs (fasting blood sugar), restcg (resting electrocardiographic measurement), thalach (maximum heart rate), exang (exercise-induced angina), oldpeak, and slope are considered in the data set.

We have designed the GUI model using Python programming, in which the user has to provide various inputs like age, male or female. Chest pain type, blood pressure, cholesterol level, fasting blood sugar, resting ECG, max heart rate, exercise angina, ST depression, and slope of ST. Depending on the input provided by the user, the random forest classifier will give a prediction, either the possibility of heart disease or no heart disease.

RESULT:

In order to assess the effectiveness of our created model, 21 inputs with varying parameter ranges were evaluated, and the outcomes were compared to the actual output. The model's output is displayed in Table No. 1 below.

Sr. No.	Age	Sex	Cp	Trestbps	Chol	Fbs	Restecg	Thalach	Exang	Oldpeak	Slope	Target Output	Achieved Output
1	24	1	2	152	150	1	2	200	1	5	2	0	0
2	25	1	3	200	500	1	2	200	1	5	1	1	1
3	29	1	1	130	204	0	0	202	0	0	2	1	1
4	35	1	0	120	198	0	1	130	1	1.6	1	0	0
5	39	0	2	138	220	0	1	152	0	0	1	1	1
6	40	1	0	152	223	0	1	181	0	0	2	0	0
7	45	0	0	138	236	0	0	152	1	0.2	1	1	1
8	46	0	2	142	177	0	0	160	1	1.4	0	1	1
9	49	0	1	134	271	0	1	162	0	0	1	1	1
10	50	1	2	140	233	0	1	163	0	0.6	1	0	0
11	55	0	1	135	250	0	0	161	0	1.4	1	1	1
12	61	0	0	130	330	0	0	169	0	0	2	0	0
13	61	1	0	120	260	0	1	140	1	3.6	1	0	0
14	62	0	0	140	394	0	0	157	0	1.2	1	1	1
15	63	0	0	124	197	0	1	136	1	0	1	0	0
16	63	1	3	145	233	1	0	150	0	2.3	0	1	1

17	64	1	2	140	335	0	1	158	0	0	2	0	0
18	65	0	2	140	417	1	0	157	0	0.8	2	1	1
19	70	1	0	145	174	0	1	125	1	2.6	0	0	0
20	70	1	2	160	269	0	1	112	1	2.9	1	0	0
21	77	1	0	125	304	0	0	162	1	0	2	0	0

Table No. 1: Comparison of Target output to Achieved output for different range of parameters.

Figure 2 shows the screenshot of the predicted output as no heart disease for input sr. no. 1 from table no. 1, and Figure 3 shows the predicted output as the possibility of heart disease for input sr. no. 2 from table no. 1.

Fig 2: Screenshot showing the predicted output as no heart disease

Fig 3: Screenshot showing the predicted output as possibility of heart disease

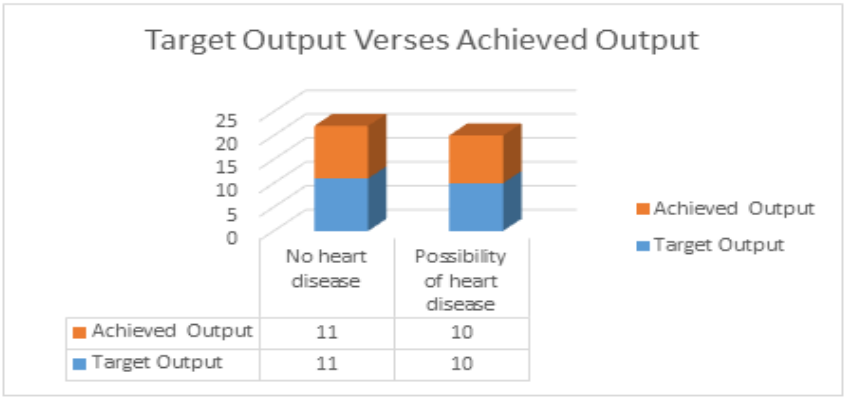

Fig 4: Graphical Representation of Target output to Achieved output.

A total of 21 inputs were tested on the model, out of which 11 were of no heart disease patience and 10 were of the possibility of heart disease patience, and our model accurately identified the same, which is shown in figure 4..

CONCLUSION

One of the newest and most popular technologies is machine learning. The application of machine learning in the healthcare sector to forecast health issues in people is growing quickly and is essential to saving lives. The Using the Python programming language, we have created a GUI-based model for predicting heart illness based on a random forest method for this study, and the results show that we are getting 100% accuracy. The output response time of the GUI-based model is quite low, and it is very simple to comprehend and use.

REFERENCES

1. Sree, P. K., Prasad, M., PBV, R. R., Ramana, C. V., Murty, P. T. S., Mallesh, A. S., & Raju, P. J. R. S. (2023, October 10). A Comprehensive Analysis on Risk Prediction of Heart Disease using Machine Learning Models. *International Journal on Recent and Innovation Trends in Computing and Communication*, *11*(11s), 605–610. https://doi.org/10.17762/ijritcc.v11i11s.8295
2. Jummelal, K. (2023, May 29). *Chronic Heart Failure Diagnosis from Heart Sounds Using Machine Learning and Full-Stack Deep Learning*. https://www.jclmm.com/index.php/journal/article/view/1065
3. Alkurdi, A. A. H. (2023, January 1). *Enhancing Heart Disease Diagnosis Using Machine Learning Classifiers*. https://doi.org/10.54216/fpa.130101
4. Swain, S. K., Behera, N., Swain, A. K., Jayasingh, S. K., Patra, K. J., Pattanayak, B. K., Mohanty, M. N., Naik, K. D., & Rout, S. (2023, October 7). *Application of IoT Framework for Prediction of Heart Disease using Machine Learning*. International Journal on Recent and Innovation Trends in Computing and Communication. https://doi.org/10.17762/ijritcc.v11i10s.7616
5. Swain, S., Chakravarty, S., Paikaray, B., & Bhoyar, H. (2023). Heart Disease Detection Using Machine Learning Techniques. *Lecture Notes in Electrical Engineering*, 273–284. https://doi.org/10.1007/978-981-19-9090-8_24
6. Zabeeulla, M., Sharma, C., & Anand, A. (2023, March 1). Early Detection of Heart Disease Using Machine Learning Approach. *CARDIOMETRY*, *26*, 342–347. https://doi.org/10.18137/cardiometry.2023.26.342347
7. Hadianti, S. (2023, September 7). Optimization of The Machine Learning Approach using Optuna in Heart Disease Prediction. *Journal Medical Informatics Technology*, 59–64. https://doi.org/10.37034/medinftech.v1i3.15
8. Nayeem, M. J. N., Rana, S., & Islam, M. R. (2022, November 30). Prediction of Heart Disease Using Machine Learning Algorithms. *European Journal of Artificial Intelligence and Machine Learning*, *1*(3), 22–26. https://doi.org/10.24018/ejai.2022.1.3.13
9. Rindhe, B. U., Ahire, N., Patil, R., Gagare, S., & Darade, M. (2021, May 12). Heart Disease Prediction Using Machine Learning. *International Journal of Advanced Research in Science, Communication and Technology*, 267–276. https://doi.org/10.48175/ijarsct-1131
10. Ruqiya. (2023, June 30). Review on Cleveland Heart Disease Dataset using Machine Learning. *Volume 21, Issue 1*, *21*(1), 87–98. https://doi.org/10.52584/qrj.2101.11
11. H, T. K. (2023, January 5). Prediction of Heart Disease Using Machine Learning with Data Mining. *Physical Science & Biophysics Journal*, *7*(1), 1–6. https://doi.org/10.23880/psbj-16000228

12. Anusuya, V., & Gomathi, V. (2021, March 25). An Efficient Technique for Disease Prediction by Using Enhanced Machine Learning Algorithms for Categorical Medical Dataset. *Information Technology and Control*, *50*(1), 102–122. https://doi.org/10.5755/j01.itc.50.1.25349

13. Online, E. (2023, September 29). *World Heart Day: 27% of deaths in India are caused by cardiovascular diseases*. The Economic Times. https://economictimes.indiatimes.com/magazines/panache/world-heart-day-27-per-cent-of-deaths-in-india-are-caused-by-cardiovascular-diseases/articleshow/104031929.cms

14. D. (n.d.). *Datasets/DS/heart.csv at main · DataThinkers/Datasets*. GitHub. https://github.com/DataThinkers/Datasets/blob/main/DS/heart.csv