Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 19 No. 16 (2023) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v19i16.42071

SHORT PAPER

Ameliorating Heart Diseases Prediction Using Machine Learning Technique for Optimal Solution

Nirmalajyothi Narisetty[1], Archana Kalidindi[1](✉), Meher Vaishnavi Bujaranpally[2], Nithyasri Arigela[2], Ch Vinod Varma[3]

[1]VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, Telangana, India

[2]Malla Reddy University, Hyderabad, Telangana, India

[3]SRKR Engineering College, Bhimavaram, Andhra Pradesh, India

archana_kalidindi@vnrvjiet.in

ABSTRACT

In the era of lacking physical fitness, folks in society are facing vital health complications which can be due to a variety of reasons such as pollution, work pressure, food, and sleeping patterns, etc. One such critical health issue which can significantly affect the regular physical activity of patients is heart disease, emulating early predictions can protect lives by inculcating proposed work and improved methodology for early heart disease prediction. In this paper, an improved methodology for heart disease prediction is proposed using a logistic regression classifier, by tuning hyperparameter using a grid-based solver with ten-fold cross-validation. The dataset used in the work is driven from the UCI Machine learning repository to evaluate the efficacy of the proposed model. Medical researchers or doctors can evaluate the model's accuracy and gets its performance. The enhanced performance rate of 90.16% has been shown in the experimental results which portrays it outclasses many prevailing models. Using this upgraded model can truly reduce the mortality rate of heart disease patients. In addition to the proposed model to make it user-friendly, a user interface has been designed where symptoms can be given as input and receives prediction as output.

KEYWORDS

disease prediction, user interface, machine learning, logistic regression

1 INTRODUCTION

The heart plays an important role in the human body. Blood is pumped through all parts of the body. If the heart fails to perform its function, it may lead to a person's death. The World Health Organization (WHO) reports that, as of 2021, about 17.9 million deaths worldwide occur from heart disease [1]. The death rate is relatively high. Several factors can contribute to heart disease, including high glucose levels, high blood pressure, smoking, and alcohol consumption. A heart attack is caused by a blockage of the coronary arteries. Whenever blood flow is reduced, red blood cells (RBCs) begin to decrease. As a consequence, the human body is deprived

Narisetty, N., Kalidindi, A., Bujaranpally, M.V., Arigela, N., Varma, C.V. (2023). Ameliorating Heart Diseases Prediction Using Machine Learning Technique for Optimal Solution. *International Journal of Online and Biomedical Engineering (iJOE)*, 19(16), pp. 156–165. https://doi.org/10.3991/ijoe.v19i16.42071

Article submitted 2023-06-07. Revision uploaded 2023-08-19. Final acceptance 2023-08-26.

of oxygen, and consciousness is lost. It is possible to prevent heart attacks through early diagnosis if the symptoms and signs are accurately predicted.

It is possible to prevent heart disease by following a healthful lifestyle and it is highly beneficial for the patient to undergo early treatment after diagnosis to maximize their chances of achieving positive results. Heart disease risks are difficult to identify when "diabetes, high blood pressure, or cholesterol" problems are present. When those situations occur, machine learning can help to diagnose heart disease at an early stage. Cardiovascular diseases can be treated using a variety of clinical methods, including drugs and medical equipment.

Nowadays, computerized techniques are used to predict cardiovascular diseases, which are far more expensive. The best way to prevent diseases is to prevent them from becoming infections. This is another way to predict cardiovascular diseases using machine learning algorithms. In the field of medical diagnosis, the use of machine learning has increased significantly, resulting in the reduction of manual error and the increase in accuracy that can be achieved through computer analysis. A subset of artificial intelligence, machine learning attempts to predict the desired results of a task by providing a well pre-processed dataset based on past results, without human intervention.

Using machine learning techniques, a disease can be diagnosed more accurately. A variety of classification algorithms are available to predict diseases such as liver disease, heart disease, diabetes, and tumors based on machine learning concepts; similarly, "regression algorithms, Random Forest, Lasso, and Logistic Regression" are used in the medical field [2], [3].

Among this, logistic regression is the most widely used statistical tool in analyzing health related data. Logistic regression can be used "to estimate the probability (or risk) of a particular outcome given the value(s) of the independent variable(s)".

An improved method for predicting heart disease through User Interface (UI) is the key objective of this research work. The implementation of this methodology enables the users to sit at a place of their convenience and check the state of their heart disease and its effects. The UI is designed in such a way that anyone can easily operate it and can predict the disease based on their symptoms.

2 LITERATURE SURVEY

Heart disease prediction has been studied for the last decade using various machine learning approaches. This section will discuss some of the approaches proposed by various researchers to analyze the several algorithms in predicting heart disease.

According to a study carried out by Abdul Saboor et al. in [4], nine machine learning classifiers were used with a grid search-based algorithm to select adequate hyperparameters to optimize the model for improving the predicted rate of heart disease. Finally, various performance metrics, including precision, recall, F-measure, and accuracy, are used to evaluate the model's performance. According to the experimental results, the best accuracy is achieved with SVM classifiers, 96.72%.

The work presented by Harshit Jindal et al. in [5] demonstrated the heart disease prediction model using three machine learning models such as "KNN, Logistic Regression, and Random Forest Classifier". The heart disease dataset was collected from the UCI repository [6]. The experiments were conducted using three different algorithms and evaluated the models based on the performance metric accuracy. Based on the experimental results it is concluded that the KNN is the highest among the three algorithms with 88.52%. However, in this study, only accuracy is considered as a performance metric.

In [2], the authors presented a machine-learning based approach to predict the heart disease adopting the Cleveland heart disease database. Several ML approaches such as classification and regression are employed, including "Random Forest, Decision Tree, and Hybrid Model". The hybrid model is a fusion of regression and classification techniques (Random Forest and decision tree) that works based on probabilities. The proposed model can predict heart disease with 88% accuracy.

The author employed a unique methodology in [7] to predict heart disease through User Interface (UI), which may be useful in everyone's data to day life. The advantage of this methodology is that a non-technical user or a person who has no programming language also can interact with the application. The model predicts heart disease based on three different algorithms as Decision Tree, Random Forest, and Logistic Regression. Users can choose any algorithm to predict heart disease. However, the authors did not discuss the performance measures of the models in this study.

An Intelligent System using machine learning has been implemented in [8]. Four different machine learning techniques are applied to detect heart attack prediction. Finally, the authors concluded that the result reveal which is the best algorithm to optimize the prediction. In contrast to traditional decision support systems, this system can answer complex queries for diagnosing heart disease, which helps healthcare professionals in making comprehensive clinical decisions. But the authors did not publish any experimental results in this study.

In [9], a web-based machine learning application is implemented. Data cleaning, pre-processing, and scaling techniques are carried out on the heart disease dataset which was downloaded from [10], followed by, the three techniques, namely, Support Vector Machine (SVM), Logistic Regression, and Naïve Bayes Algorithm. These were trained with 75% dataset for training and the remaining dataset for testing, respectively. Based on the experimental results, it is observed that the SVM classifier yields the best accuracy among the three with 64.4%. Therefore, SVM was identified as the most effective algorithm for the web application. However, the accuracy can be improved by integrating principal component analysis into the model.

In [11] the authors have used the powerful evolutionary algorithm Particle Swarm Optimization (PSO). The PSO technique works based on the rule mining concept. The author collected datasets from the UCI repository of Machine Learning Databases. A comparison is made between the accuracy results and the decision tree based on the C4.5 algorithm. According to the experiment, the PSO algorithm achieved a high predictive accuracy of 87% and outperformed C4.5 by a wide margin (63% to 73%). The error rate of the proposed method is 13%, which is less when compared to the error rate of the decision tree (19%). However, the authors in this study did not consider the computational complexity.

A machine learning-based heart attack prediction method (ML-HAP) was presented by Nandal N et al. in [12], which used various machine learning algorithms, including "Support Vector Machines, Logistic Regression, Naïve Bayes, and XGBoost". The researchers used various parameters in this study to enhance the prediction. Out of the four algorithms, XGBoost has the highest prediction with 92 and logistic regression has 85. As compared to [12], [13] reports the work of Hassan C. et al. They have used eleven machine learning classifiers to identify key features to improve heart disease prediction rate. A combination of these features and well-known classification algorithms was used to introduce the prediction model. One of them is logistic regression, which achieved a prediction rate of 88.25%.

The studies cited above mainly focused on comparing the performances of various machine learning classifiers. The time complexity is also one of the important metrics in evaluating the classifier which was not considered. Any non-technical

person may not be able to use it if he or she may want to confirm the disease based on their heart disease symptoms or test results. To fulfill these research gaps, the present study is carried out for improving the performance of the ML classifier and which is connected with advanced technologies such as flask interface for providing UI so that users can easily operate it.

3 METHODOLOGY

Every day, there are more number cases of heart disease diagnosed. This is unfortunate. So, an accurate risk assessment system was necessary, based on demographic information, medical history, physical examination, behavioral analysis, and laboratory tests. A step-by-step procedure must be followed in machine learning algorithms to improve accuracy. Exploratory data analysis has been performed on the dataset at the very prior step followed by understanding the behavior of each variable using visualization. The proposed methodology consists of the following stages: i) Data collection ii) Pre-processing and iii) Attribute standardization and iv) Logistic Regression, as shown in Figure 1.

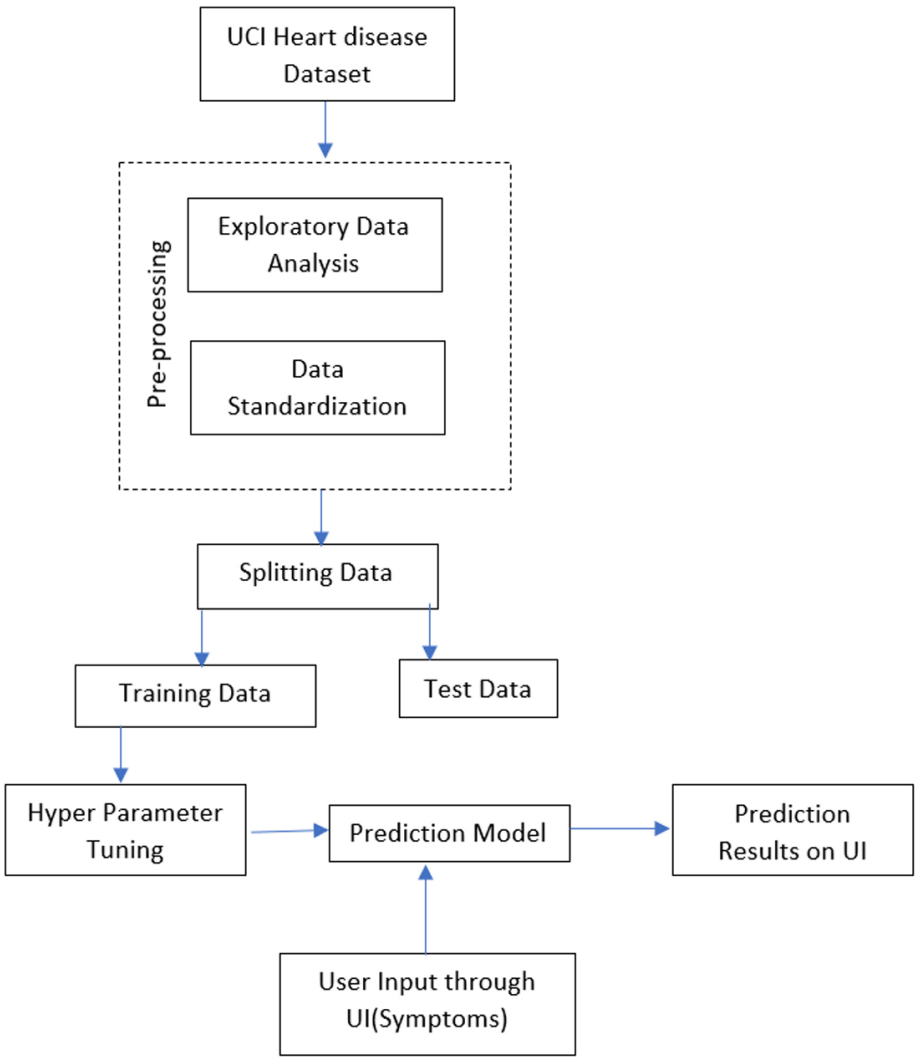

Fig. 1. Proposed methodology for heart disease prediction

3.1 Data collection

The heart disease dataset is collected from the UCI machine learning library [6]. The dataset consists of 14 attributes and 303 records. As shown in the last column, the target value is either zero or one depending on the existence or absence of disease in the body of the patient.

3.2 Exploratory data analysis

It is an important step in preparing the data which is input to the ML classifier. As part of exploratory data analysis, it was checked for null values, duplicate rows, and outliers. Null values are replaced with mean values, one duplicate row, and dropped; outliers are identified by plotting the boxplots for numeric attributes based on minimum, maximum, and Inter Quartile Range (IQR) values. Outliers are identified in 'trtbps', 'chol',' oldpeak' attributes and handled properly. Then the categorical attributes are converted into numeric by introducing the dummy variables. As a result of dataset, all the categorical attributes and numerical attributes are aggregated.

3.3 Attribute standardization

It is an important and necessary activity in the field of data analysis. The attribute values may influence the data analysis or classification process. For example, attributes with higher values can dominate the attributes with less value.

The qualitative attributes are converted in the transformation step. Now the attributes need to be normalized to eliminate such dominance by scaling them all within a specific range based on the standard deviation and mean value of that attribute [14], [15]. In this paper for the normalization process, the Z-score normalization technique is applied using equation (1).

$$z = \frac{X - \mu}{\sigma} \tag{1}$$

3.4 Logistic regression and hyperparameter tuning

This section discusses the logistic regression for binary classification. It is one of the popular supervised learning algorithms considered in the medical field. It uses the sigmoid function as shown in equation (2) which returns the probabilistic values (0 to 1) as outcomes. This model provides the probabilistic values as outcomes.

$$f(z) = \frac{1}{1 + e^{-z}} \tag{2}$$

The resulting dataset from the previous dataset is divided into train and test sets with the ratio 80:20. Hyperparameter tuning is an optimization technique that makes the machine learning model meet the maximum performance with generalization capability. The focus has been on optimizing heart disease prediction based on solver hyperparameter. The hyperparameter "solver" by values 'newton-cg', 'lbfgs', 'liblinear', 'sag', 'saga' and "max_iter" by values 20, 50, 100, 200, 500, 1000 are tuned with the adoption of the grid-search-based algorithm. The outcome of the experiment is

solver = 'liblinear', max_iter = 20. The logistics regression is trained with the training dataset and tested on the test dataset with the identified parameter values.

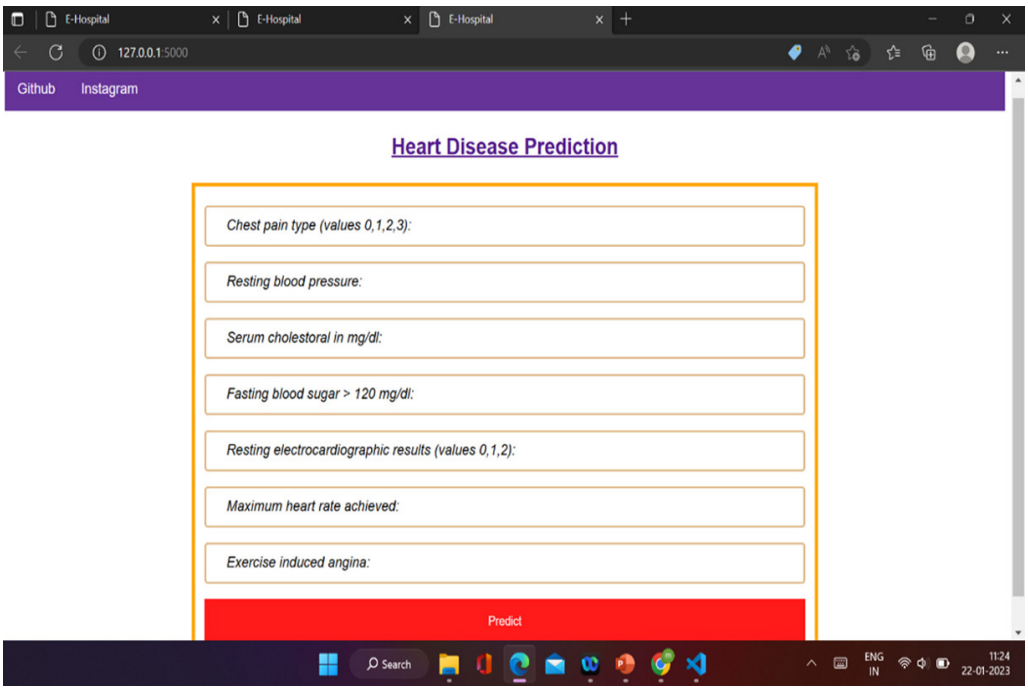

Fig. 2. Basic GUI for heart disease prediction

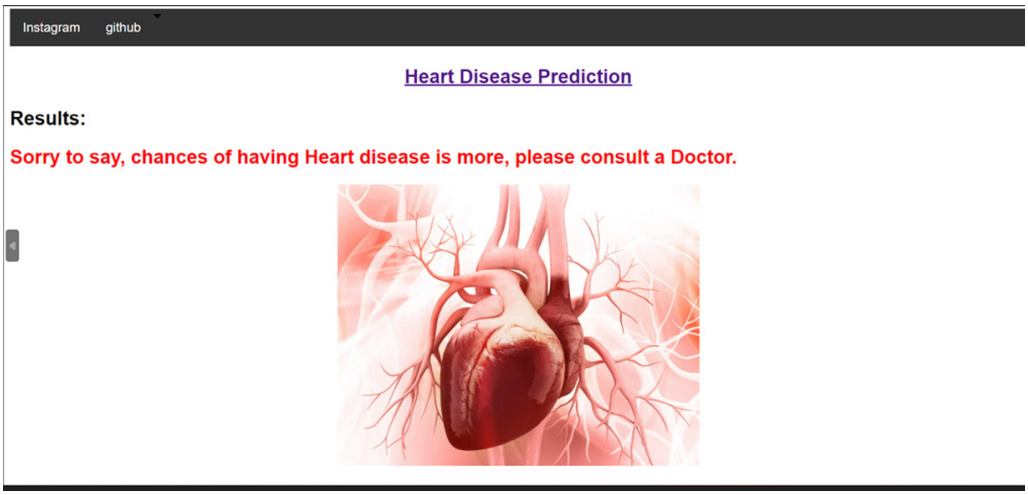

Fig. 3. Positive case of heart disease prediction

After building the model, the flask python web server gateway framework is used to display the predictions of the ML model to the user interface on a browser page. Using this framework, we can run our application in localhost and the same can be also hosted into a public web address using ngrok. It is the fastest and free service that enables you to share a webpage or server running on your local machine without deploying. Now this website address can be accessed by anyone who knows it whether on mobile/laptop from everywhere. The following Figures 2–4 shows the respective screenshots of UI for heart disease prediction for both positive and negative results.

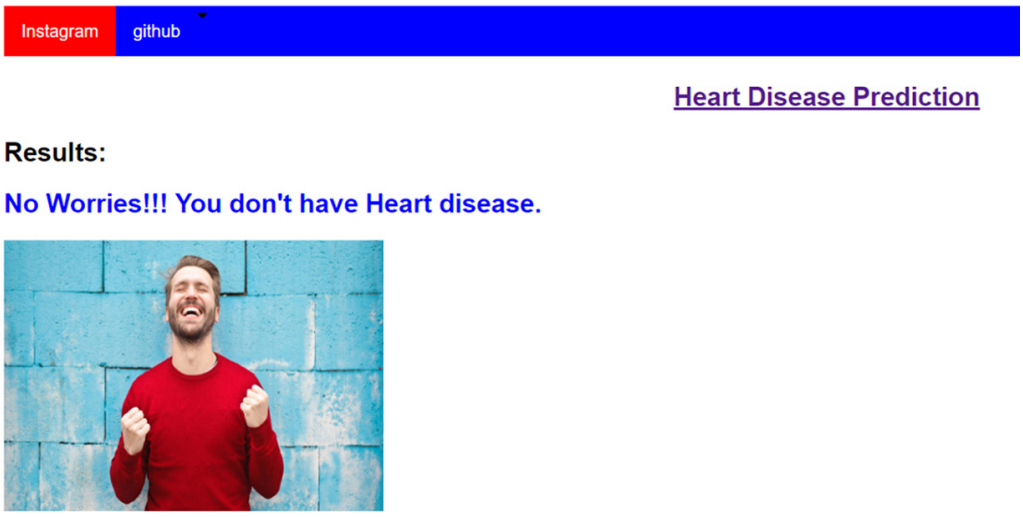

Fig. 4. Negative case of heart disease prediction

4 RESULTS AND DISCUSSION

This section discusses the experimental results. The hyperparameter tuning is a form of regularization. Further, the logistic regression classifier metrics was compared with and without hyperparameter tuning to verify the effectiveness. For evaluation purposes, the performance results are compared with basic logistic performance and shown in Table 1 as well as plotted in Figures 5 and 6.

From the literature, it is observed that several algorithms are being used by researchers in order to find patients with heart disease. SVC, Decision Trees, KNNs, Random Forest classifiers, and Logistic Regression are some of the most popular. From Table 1 it is observed that the accuracy of the logistic regression model in this study is 86.88%, which is high compared to studies [12], [13]. Further, the hyperparameter tuning contributed to enhancing the performance of heart disease and achieved 90.16% in binary classification.

Table 1. Performance metrics of for state-of-the-art approaches and proposed methodology

	Accuracy	Precision	Recall	F-Measure
Logistic Model [13]	85	83	85	82
Logistic Model [12]	88.25	87.91	88.25	88.65
Logistic Model	86.88	88.57	88.57	88.57
Proposed Model	90.16	91.17	93.93	88.57

To verify its efficacy, an evaluation has been made between logistic regression classifier before and after hyper parameter tuning. Heightened accuracy of 90.16% is achieved by the model, superior studies [12] and [13]. Effective visualization plays a major role in improving model performance, which are compared. 91.7% and 93.93% are substantial results with respect to precision and recall. However, the F-score remains the same. The values emphasize the advantages of logistic regression model supported by regularization in classification of heart diseases. The role of data analysis and visualization is evident. It would be beneficial to use the proposed model

through UI as a complementary tool to aid in diagnosing and predicting heart disease in clinical practice. Yet further research should address potential margins and plunge deeper into the unchanged F-score which can be improved.

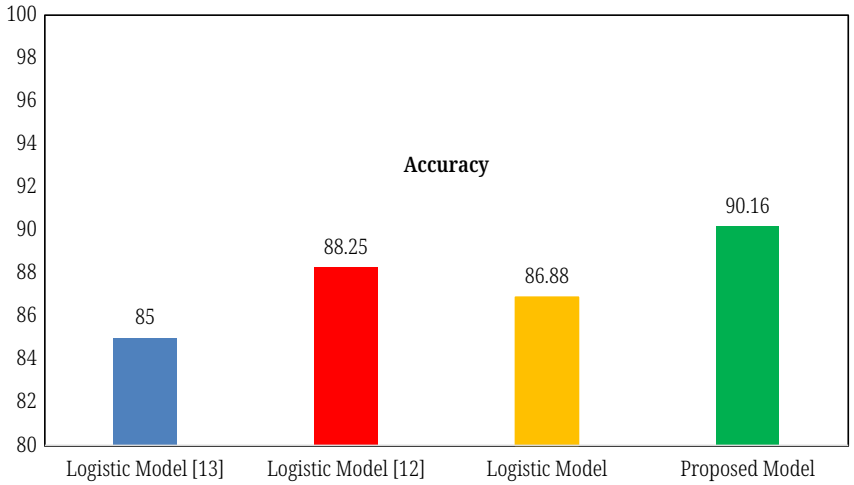

Fig. 5. Heart disease prediction through logistic regression models

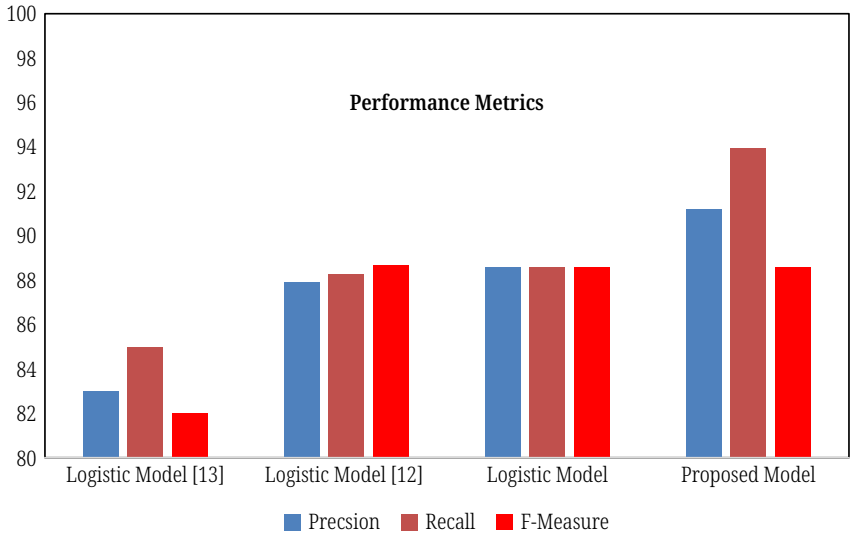

Fig. 6. Comparison of the existing models and proposed model on the basis of performance metrics

5 CONCLUSION

An improved methodology for detecting cardiovascular disease has been developed using logistic regression. The proposed methodology assists even a non-technical patient to confirm about disease based on their heart disease symptoms or clinical test results by using computer-aided techniques. The proposed model achieved accuracy is 90.16% which shows that there are higher chances of the model to accurately predict whether the given person has heart disease or not. In terms of classification accuracy, precision rate, and recall rate, experimental results show that the proposed method is capable of providing satisfactory classification performance. However, this study was focused on only one machine learning

algorithm and did not explore to the comparison of other supervised learning algorithms. Additionally, the dataset used in this study was obtained from the UCI repository which may not be representative of the general population. Future research will focus on addressing all these limitations by studying the comparative analysis of different supervised machine learning algorithms, by focusing on hybrid models to enhance prediction rate and time complexity and also to improve the f-score values. Moreover, the developed web application can be deployed into a cloud-based environment to be accessed anywhere in the world.

6 REFERENCES

[1] "Cardiovascular diseases (CVDs)," World health organization.int. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds). [Accessed June 6, 2023].

[2] K. Oqaidi, S. Aouhassi, and K. Mansouri, "Towards a students' dropout prediction model in higher education institutions using machine learning algorithms," *International Journal of Emerging Technologies in Learning*, vol. 17, no. 18, pp. 103–117, 2022. https://doi.org/10.3991/ijet.v17i18.25567

[3] A. Kalidindi, P. L. Kompalli, S. Bandi, and S. R. R. Anugu, "CT image classification of human brain using deep learning," *International Journal of Online and Biomedical Engineering*, vol. 17, no. 1, pp. 51–62, 2021. https://doi.org/10.3991/ijoe.v17i01.18565

[4] A. Saboor, M. Usman, S. Ali, A. Samad, M. F. Abrar, and N. Ullah, "A method for improving prediction of human heart disease using machine learning algorithms," *Mobile Information Systems*, vol. 2022, pp. 1–9, 2022. https://doi.org/10.1155/2022/1410169

[5] H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, "Heart disease prediction using machine learning algorithms," *IOP Conference Series: Materials Science and Engineering*, vol. 1022, p. 012072, 2021. https://doi.org/10.1088/1757-899X/1022/1/012072

[6] "UCI Machine Learning Repository: Data Set," Uci.edu. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/heart+disease(Dataset). [Accessed June 6, 2023].

[7] Bharathi, "Heart disease prediction using machine learning," *Revista GEINTEC – Gestão, Inovação e Tecnologias*, vol. 11, no. 4, pp. 3694–3702, 2021. https://doi.org/10.47059/revistageintec.v11i4.2400

[8] S. Karpagam, M. Kaleeswari, K. Kavitha, and S. Priyadarsini, "Heart disease prediction using machine learning algorithm," *International Journal of Scientific Development and Research*, vol. 5, no. 8, pp. 334–337, 2020. https://www.ijsdr.org/papers/IJSDR2008043.pdf

[9] A. Jagtap, P. Malewadkar, O. Baswat, and H. Rambade, "Heart disease prediction using machine learning," *International Journal of Research in Engineering*, vol. 2, no. 2, pp. 352–355, 2019. https://www.ijresm.com/Vol.2_2019/Vol2_Iss2_February19/IJRESM_V2_I2_89.pdf

[10] M. R. K. Sony, "UCI Heart Disease Data." 23-Sep-2020.

[11] A. H. Alkeshuosh, M. Z. Moghadam, I. A. Mansoori, and M. Abdar, "Using PSO algorithm for producing best rules in diagnosis of heart disease," in *2017 International Conference on Computer and Applications (ICCA)*, 2017. https://doi.org/10.1109/ICICT50816.2021.9358597

[12] N. Nandal, L. Goel, and R. Tanwar, "Machine learning-based heart attack prediction: A symptomatic heart attack prediction method and exploratory analysis," *F1000Research*, vol. 11, p. 1126, 2022. https://doi.org/10.12688/f1000research.123776.1

[13] C. A. U. Hassan *et al.*, "Effectively predicting the presence of coronary heart disease using machine learning classifiers," *Sensors*, vol. 22, no. 19, p. 7227, 2022. https://doi.org/10.3390/s22197227

[14] Z. Zeng, W. Peng, and B. Zhao, "Improving the accuracy of network intrusion detection with causal machine learning," *Security and Communication Networks*, vol. 2021, pp. 1–18, 2021. https://doi.org/10.1155/2021/8986243

[15] P. Nuankaew, P. Nasa-Ngium, T. Kunasit, and W. S. Nuankaew, "Implementation of data analytics and machine learning in Thailand education sector," *International Journal of Emerging Technologies in Learning*, vol. 18, no. 5, pp. 175–191, 2023. https://doi.org/10.3991/ijet.v18i05.36871

7 AUTHORS

Dr. Nirmalajyothi Narisetty earned her Doctorate in Computer Science and Engineering from Acharya Nagarjuna University, Guntur, Andhra Pradesh. She is currently working as Assistant Professor in the department of CSE-AIML & IOT, VNR Vignana Jyothi Institute of Technology, Hyderabad. She has 13 years of Teaching Experience and 3.5 years of Research Experience. Published 12 research publications in reputed international journals. Guided 10 UG projects. She published 1 Indian Patent. Her research interests include Data Mining, Machine Learning, Cloud Computing etc. (E-mail: nirmalajyothi_n@vnrvjiet.in).

Archana Kalidindi is the Assistant Professor in CSE-AIMT & IoT, VNR VJIET Hyderabad, India. She received the Bachelor's Degree in Information Technology (2007), Master's Degree in Software Engineering (2011) and currently pursuing Ph.D. in Computer Science in the area of IoT and Deep Learning. Her research interest includes Deep Learning, Data Science, Machine Learning. Published many research articles in renowned journals and conferences (E-mail: archana_kalidindi@vnrvjiet.in).

Meher Vaishnavi Bujaranpally B-Tech IV year, Department of Computer Science and Engineering, Malla Reddy University (E-mail: bmehervaishnavi@gmail.com).

Nithyasri Arigela B-Tech IV year, Department of Computer Science and Engineering, Malla Reddy University (E-mail: nithyasriarigela@gmail.com).

Ch Vinod Varma is the Assistant Professor in CSE SRKR Engineering College, Bhimavaram, Andhra Pradesh. He received the Bachelor`s Degree in CSE (2013), Master's Degree in CST (2015) and currently pursuing Ph.D. in computer science in the area of machine learning. His research interest includes Deep Learning, Data Science. Published many research articles in renowned journals and conferences (E-mail: vinodvarmaaa@gmail.com).