International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

AI-Based Prediction of Myocardial Infarction in Patients Using Various Algorithms

[1]Shridevi K. Jamage, [2]Dr. Ramesh Y. Mali, [3]Dr. Virendra V. Shete

Submitted: 23/12/2023 **Revised**: 29/01/2024 **Accepted**: 07/02/2024

Abstract: Myocardial Infarction (MI) is among the primary causes of mortality worldwide and early detection of this condition can improve patient outcomes. Artificial Intelligence (AI) have shown promise in predicting Myocardial Infarction in patients, but the optimal algorithm for this task is not yet clear. This study assessed the efficacy of four ML algorithms - K-nearest neighbours (KNN), Logistic regression, Support Vector Machine (SVM), and random forest analysis - in predicting MI in patients. This study includes Myocardial Infarction dataset of 303 patients with details of medical history, demographic info, as well as clinical constraints. The data pre-processing was done with missing values, removing outliers and normalizing the data. In addition, feature selection approaches identify the most relevant predictors of myocardial infarction. The accuracy metrics are determined by evaluating the training performance of the four algorithms on a practice set. When the results are compared, Logistic Regression outplays the others with an overall accuracy of 81.32%. However, K-nearest neighbors, SVM, and Random Forest had accuracy rates of 65.93%, 54.95%, and 81.32%, respectively. Thus, according to our research findings, Logistic Regression is the optimal algorithm for predicting MI in patients. It is a straightforward, interpretable, and efficient technique that can be used in clinical decision-making. Our findings provide essential data about the use of machine learning algorithms to predict myocardial infarction and can help guide future studies in this area.

Keywords: *Logistic Regression, Myocardial Infarction, Support Vector Machine, Prediction, K-Nearest Neighbours, Artificial Intelligence, Clinical Decision Making Random Forest Classifier*

1. Introduction

Myocardial Infarction (MI) is known as a serious and enduring disease induced by the inability of heart to circulate adequate blood to achieve the body's requirements [1]. It is a substantial concern related to global public health and the main source of sickness and mortality [2]. The occurrence of MI has been increasing owing to ageing populations, increasing the rates of hypertension and diabetes, as well as improved survival rates for other chronic diseases [3], [4].

MI is linked to a number of risk factors [5], [6], such as coronary artery disease, hypertension, diabetes, obesity and smoking [7]–[9]. Other factors such as genetics [10], viral infections [11], [12], and alcohol abuse [13] can also contribute to the development of MI. Myocardial Infarction is often accompanied by symptoms such as tiredness, loss of breath, leg swelling, and reduced exercise tolerance [14].

The diagnosis of Myocardial Infarction includes a amalgamation of clinical evaluation, imaging tests for example echocardiography, and laboratory tests to evaluate heart function. MI treatment normally involves lifestyle reforms such as exercise, dietary changes, and smoking cessation, as well as pharmacological interventions like diuretics, beta-blockers, and ACE inhibitors. In serious cases, medical interventions like ventricular assist devices or heart transplants may be necessary.

Many factors can increase the risk of Myocardial Infarction including following some of the most common risks:

- Age: The danger of developing Myocardial Infarction rises as people get older, particularly after the age of 65.
- Gender: Women are generally at a lower risk of Myocardial Infarction than men, but this changes after menopause, when their risk increases to match that of men.
- Family history: The risk MI increases with a history of MI in the family.
- High cholesterol: The chance of having Myocardial Infarction increase with an excessive amount of LDL cholesterol that result in the accumulation of plaque in the arteries.

[1]*Assistant Professor, Department of Electronics & Telecommunication Engineering, SCTR's Pune Institute of Computer Technology, Pune (MS), India, shridevikjamage@gmail.com*
[2]*Professor and Head of the Department of Electrical and Electronic Engineering, MIT School of Computing, MIT ADT University, Pune (MS), India, ramesh.mali@mituniversity.edu.in*
[3]*Professor & Director, MIT School of Engineering and Science, MIT ADT University, Pune (MS), India, virendra.shete@mituniversity.edu.in*

- High blood pressure: High BP can harm the arteries which may further boost the chance of emerging Myocardial Infarction disease.
- Diabetes: It may lead to damage in the blood vessels, which then raises the likelihood of atherosclerosis. Then, it may increase the risk of Myocardial Infarction.
- Obesity: In obesity, an excessive amount of weight can put stress on the heart. It causes increase in the risk of diabetes, hypertension, and high cholesterol. All these can lead to Myocardial Infarction.
- Smoking: The tobacco consumption harm the blood vessels and increase the chance of developing Myocardial Infarction disease.
- Dearth of physical movement: Insufficient exercising can raise the probability of developing Myocardial Infarction.
- Stress: The stress for long time increase blood pressure and inflammation in the body that may lead to Myocardial Infarction.

These factors can interrelate with each other, besides multiple risk factors can greatly strengthen the probability of developing MI.

It is important to detect MI early and its management to prevent disease progression and enhance outcomes. ML algorithms in forecasting MI in patients displays considerable potential, offering a latent tool for early detection and intervention. However, a study needs to refine and enhance the precision as well as effectiveness of algorithms for clinical use. Invasive techniques for MI Diagnosis involves analyzing the symptoms, results of physical examination with medical history. However, they are prone to errors resulting in an imprecise diagnosis, and delays in treatment. Moreover, the techniques are frequently computationally complex, costly, as well as time-consuming. Non-invasive MI diagnosing techniques such as stress test, ECG, cardiac MRI (magnetic resonance imaging), and echocardiography, are increasingly popular due to effectiveness and lower risk.

ECG is a popular for assessing heart function. It detects anomalies in the electrical activity of the heart. Echocardiography (Echo Test) uses sound waves to create pictures of the heart. It describes the structure and function of the heart. Cardiac MRI creates comprehensive pictures of the heart by employing magnetic fields and radio waves, whereas stress testing monitors the heart's reaction to exercise or treatment.

These non-invasive techniques are effective in diagnosing MI disease and have the advantage of being safer and less expensive than invasive techniques such as cardiac catheterization. However, they do have limitations, such as the need for specialized equipment and expertise, and the inability to provide detailed information about the heart's blood vessels.

ML algorithms have the potential to increase the accuracy and efficiency of non-invasive diagnostic procedures for Myocardial Infarction illness. ML algorithms can uncover patterns and predictors of MI illness and aid physicians in establishing more accurate diagnosis by evaluating big datasets of patient information. These algorithms can also assist to decrease the risk of human mistake and increase diagnosis speed, resulting in early intervention and improved patient outcomes.

Researchers developed a non-invasive medical decision support system that combines AI prediction models since invasive procedures for detecting MI can be challenging. Logistic regression (LR), decision tree (DT), SVM, fuzzy logic (FL), artificial neural network (ANN), Naive Bayes (NB), AdaBoost (AB), k-nearest neighbour (K-NN), and rough set are among the models included.

These machine-learning-based medical decision-making systems have been widely employed for MI diagnosis, resulting in a reduction in the ratio of MI disease-related mortality. These tools help medical personnel make more accurate diagnosis and treatment decisions by analysing massive datasets and discovering patterns and predictors of cardiac disease. This has led in early intervention and better results for MI disease patients.

2. Literature Review

Youshen Xia and Jun Wang compared the performance of a two-layer neural network and a SVM for classification tasks. They discovered that the neural network algorithm not only had a lower implementation complexity, but it was also able to converge faster to the optimal solution of SVM learning, achieving an accuracy of 84.1% on a test set [15]. Monika Gandhi and Shailendra Shigh employed Data mining algorithms like Neural network, NB, and decision tree to forecast MI in patients [16]. According to the Trivandrum MI Registry, hospitalization in India because of MI is the most typical cause of cardiac-related illness. The registry showed that 59% of those admitted to the hospital with MI passed away within 5 years of being monitored [17]. Nahar et al. [18] applied three rule generation algorithms - Tertius, Predictive Apriori, and Apriori - for rule mining and found that women have a lower probability of coronary MIthan men. They also identified characteristics that can show if a person is healthy or ill. The results of the research showed that both males and females are prone to developing MIif they experience chest pain without any symptoms and angina that is brought on by exercise. Ali et al. presented two

SVM models in the expert system that increased the prediction accuracy of MI by 3.3% compared to the standard SVM model [19]. Udovychenko and Illya assessed 800 current density distribution maps (CDDM) using the kNN classifier method to classify the heart state groups. Their analysis achieved an accuracy ranging from 80-88%, with a sensitivity of 70-95% and precision of 77-93% [20].

3. Methodology

3.1Dataset Description

The UCI AI repository provides access to the Cleveland MI Data Set. [21], is a multivariate dataset that contains 303 instances with 75 attributes. This dataset has been used extensively for research purposes in the domain of MI prediction using ML algorithms.

Out of the 75 attributes, only 14 attributes are used for MI prediction, which is as follows:

1. 3 (age): The age of the patient in years as a numeric value.
2. 4 (sex): A patient's gender is denoted as 1 for male and 0 for female.
3. 9 (cp): The patient's chest pain.
4. 10 (trestbps): Patient's blood pressure at his rest state.
5. 12 (chol): Cholesterol concentration in the serum.
6. 16 (fbs): Glucose level after a fast.
7. 19 (restecg): ECG results.
8. 32 (thalach): The highest rate of heart attained by the patient.
9. 38 (exang): angina induced by exercise- is indicated as 1 for affirmative and 0 for negative.
10. 40 (oldpeak): The amount of ST depression brought on by physical activity compared to when at ease.
11. 41 (slope): The ST segment's slope at the highest level during exercise.
12. 44 (ca): The count of significant blood vessels (ranging from 0 to 3) that can be visualized through fluoroscopy is counted.
13. 51 (thal): The outcome of the Thallium stress test.
14. 58 (num): A MI diagnosis (status of angiographic illness) is indicated by 0 for absence and 1, 2, 3, or 4 for presence.

By examining the provided information, this dataset aims to establish whether or not a person has MI disease. Attribute #58 (num) is the predicted attribute, which is used as the target variable for the MI prediction model.

The dataset is publicly available and has been extensively used in research studies to evaluate and compare the effectiveness of various ML algorithms for MI prediction. The dataset preprocessing and cleaning make it suitable for direct use in machine-learning experiments. The dataset can be used to develop and assess a range of supervised AI methods, such as SVM, DT, LR, and ANN etc.

3.2Proposed methods

In proposed method, 70% of the data is assigned to the training set, while the remaining 30% is assigned to the testing set. AI models are created by utilizing the training set, and then their accuracy is determined by using the testing set.

The following libraries are used in the program:

1. pandas - Used for manipulating and analyzing data.
2. numpy - Used for computation with numbers.
3. matplotlib - Used for data visualization.
4. seaborn - Used for data visualization.
5. sklearn - Used for AI.

The supervised AI technique predict MI disease. It uses classification models to determine whether a patient suffers from MI illness or not according to several medical and demographic features.

The program defines several functions for the data preparation and model performance assessment:

1. load_data() - Loads the dataset into a pandas data frame.
2. preprocess_data() - Preprocesses the data by removing missing values, encoding categorical variables, and scaling the features.
3. plot_feature_importance() - Plots the importance of each feature in the dataset.
4. plot_confusion_matrix() - Plots the confusion matrix for a given model.
5. evaluate_model() - Assesses how well a model performs on the test set.

The program uses the following classification models to predict MI disease:

1. Logistic Regression - A linear regression model that utilizes a sigmoid function to forecast the probability of a binary outcome.
2. KNN – It is non-parametric model that assigns the class of a given data point based on the k-closest data points in the training set.
3. SVM – It is a parametric model that identifies the optimal hyperplane to differentiate between the classes in the training data.
4. Random Forest - An ensemble model that combines multiple decision trees to improve the accuracy of predictions.

The following metrics are considered to evaluate the testing performance of each model.

1. Accuracy - The fraction of correct predictions.

2. Precision - A measure of how many positive predictions were correct.
3. Recall - A measure of how many positive instances were correct.
4. F1_score - The harmonic average of precision and recall.
5. Support - The number of instances in each class.

The training performance of each model determines if there is any over fitting. The accuracy calculate the percentage of accurate forecasts made by a model on the test set. The model's accuracy is calculated by dividing the total number of accurate predictions by the total number of predictions produced.

The test performance of the models is evaluated using precision, recall, f1-score, and support. It is based on the confusion matrix and provides a summary of the model's predictions in terms of true and false positives, true and false negatives.

The precision is obtained by dividing the number of true positive predictions to the total number of positive predictions including true and false positives. The equation for precision of a model is shjown below.

$$Precision = \frac{TP}{(TP+FP)} \quad (1)$$

However, recall includes the number of correct positive predictions divided by the total actual positives for correct and incorrect predictions as shown in the formula below.

$$\text{Recall} \frac{TP}{(TP+FN)} \quad (2)$$

The F1-score based on precision and recall is widely utilized to provide a balanced assessment. It combines the harmonic average of precision and recall and is calculated as $\frac{2\,x\,(precision\ x\ recall)}{(precision + recall)}$.

The metric support is the number of instances in each class. On the other hand, a confusion matrix in AI, is a table-based illustration summarizing the effectiveness of a classification model. It is obtained by applying a test dataset with known true labels and gives four values that provide information on the model's performance.

- True Positive (TP): It represents the count of instances that the model correctly categorized as belonging to the positive class.
- False Negative (FN): It shows the number of instances that are actually positive but are incorrectly classified as negative by the model or the number of times the model fails to identify a positive instance.
- False Positive (FP): It gives the number of instances belong to the negative class that the model incorrectly tags as positive.
- True Negative (TN): It shows how many negative cases the model has correctly identified.

The performance of a classification model can be assessed through a confusion matrix, which can calculate key metrics like precision, accuracy, recall, and F1-score. This matrix provides a visual representation of the model's outcomes, enabling the identification of potential errors.

Figure 1 displays the flowchart for the AI technique for MI forecasts.

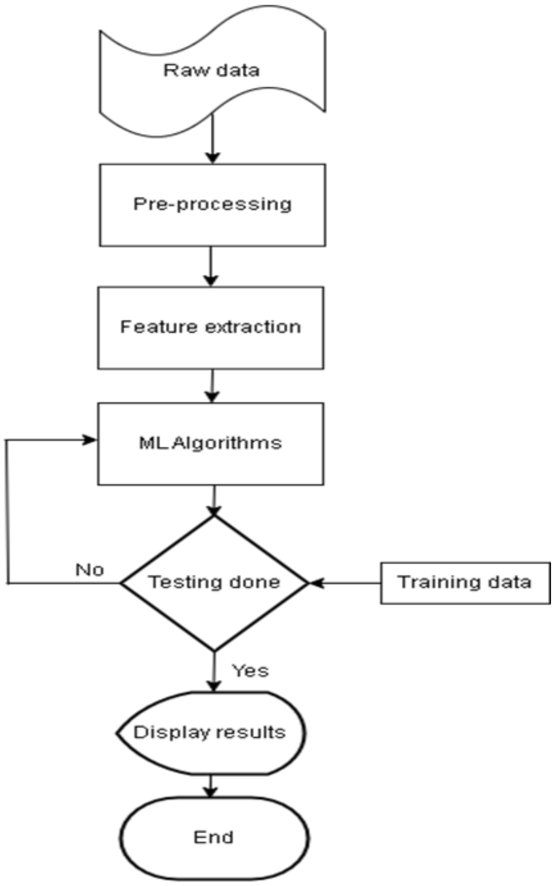

Fig 1 Flow chart for MI forecasts using AI

3.3 Algorithms employed in the recommended system:

3.2.1Logistic Regression:

Logistic regression, an analytical statistical approach, has gained significance in the realm of artificial intelligence. It leverages current data to predict outcomes with two alternatives, enabling machine learning systems to categorize incoming data by comparing it to historical data. As more information is incorporated, these algorithms enhance their predictive capabilities. Logistic regression plays a crucial role in data preparation for evaluation, where the Extract, Transform, Load (ETL) procedure is employed to classify data into predefined categories. The logistic regression model is formulated using a logistic function, denoted as f(z):

$$f(z) = \frac{1}{1+e^{z}} \quad (3)$$

Here, the variable z takes on values such that the output range of f(z) always falls between 0 and 1. This mathematical formulation underlies the logistic regression approach, providing a framework for predictive modeling and classification in machine learning [22].

3.2.2KNN:

Belonging to the supervised class of machine learning algorithms, k-Nearest Neighbours (kNN) relies on labelled training data for generating predictions. However, when the objective is not prediction but rather the identification of similar data points, kNN can be effectively applied in an unsupervised manner. It's crucial to emphasize that kNN is a highly versatile algorithm capable of addressing a wide range of problems.

$$J(v) = \sum_{i=1}^{C} \sum_{j=1}^{Ci} (||Xi - Vj||)^2 \qquad (4)$$

The equation represents the count of data points within the ith cluster, quantified by the Euclidean distance ‖Xi-Vj‖. Referring to the Euclidean distance, this expression pertains to the distance measurement between the data points Xi and Vj and the center of the cluster [23].

3.2.3SVM:

The SVM (Support Vector Machine), a frequently utilized supervised learning algorithm, is applied to address both Classification and Regression problems. While its applications are diverse, the conventional SVM model is predominantly employed in artificial intelligence for handling classification challenges. The formula representing this model is as follows:

$$f(x) = \sum_{i=1}^{n} (\propto_i y_i K(x_i, x) + b) \qquad (5)$$

In this context, the expression involves the kernel function denoted as K(xi, x), incorporating the ith sample xi. The parameters to be estimated are represented by αi and b [15].

3.2.4Random Forest:

The Random Forest, an artificial intelligence technique, involves constructing multiple decision trees during the training phase. By utilizing samples from the training dataset and selecting different attributes for each tree, the algorithm generates a diverse set of decision trees. This intentional introduction of randomness and variability aims to mitigate overfitting and enhance the overall generalization performance of the model. In the case of regression tasks, the algorithm computes and returns the average prediction from each tree. The Random Forest methodology covers individual classification with regression tree classifiers, as illustrated in the equation [24].

$$\{h(x, \theta k), k = 1,2, \dots i \dots\} \qquad (6)$$

In this equation, the input variable is denoted as x, θ_k represents a randomly selected predictor variable, and h indicates the Random Forest classifier.

4. Result

Four unique AI models— K-Nearest Neighbors, Logistic Regression, Support Vector and Random Forest Classifier—were tested after analyzing the Cleveland MI Data Set.

1. Logistic Regression Model: The Logistic Regression model obtains an accuracy of 87.26% on the training set and 81.32% on the test set. Notably, the model excels at predicting individuals without cardiac disease, with a precision of 0.89 in the training set and 0.80 in the test set. The model has a training and a test set recall score of 0.91 and 0.84 for detecting individuals with MI.
2. K-Nearest Neighbors Classifier: The accuracy of the K-Nearest Neighbors Classifier is 76.89% on the training set and 65.93% on the test set.
3. Support Vector Classifier: While the Support Vector Classifier achieves a perfect accuracy of 100% on the training set, it achieves a much lesser accuracy of 54.95% on the test set. This disparity indicates a lack of generalization and probable over fitting to the training data.
4. Random Forest Classifier: In the Cleveland Heart Disease Data Set, the Random Forest Classifier displays significant prediction performance for MI. It achieves 100% training set accuracy and 81.32% test set accuracy, demonstrating strong performance. Furthermore, the model receives a score of 1 on the training set and a score of 0.79 on the test set.

The Logistic Regression model stands out as the top performer, boasting the uppermost training and testing sets accuracy. Conversely, the SVM exhibits over fitting tendencies, making it less suitable for this particular data set. The K-Nearest Neighbours performs less effectively as compared to Logistic Regression for MI prediction in patients.

Table 1 AI Model with different parameters

Model	Precision		Recall		F1-Score		Support	Training Accuracy %	Testing Accuracy %
	MI No	MI Yes	MI No	MI Yes	MI No	MI Yes	No. of Instances		

Logistic Regression	0.8	0.82	0.78	0.84	0.79	0.83	41	50	87.26	81.32
K-NN	0.63	0.68	0.59	0.72	0.61	0.7	41	50	76.89	65.93
Support Vector Machine	0	0.55	0	1	0	0.71	41	50	100	54.95
Random Forest Classifier	0.8	0.82	0.78	0.84	0.79	0.83	41	50	100	81.32

Table 1 demonstrates names and parameters of models.

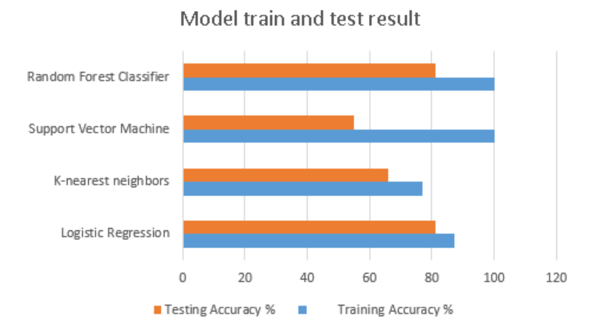

Fig 2. Training and testing results

In Figure 2, a graph represents the results obtained by the training and testing sets.

5. Discussion

This program utilizes supervised machine learning algorithms to predict the presence of myocardial infarction (MI) using the Cleveland Heart Disease dataset. With 303 instances and 75 attributes, the dataset allocates 70% to the training set for constructing AI models, while the remaining 30% serves as the testing set for accuracy evaluation.

Various classification models, including Logistic Regression, k-Neighbours Classifier, Support Vector Machine (SVM), and Random Forest Classifier, are employed to predict MI. Data pre-processing involves handling missing values, encoding categorical variables, and scaling features.

The program assesses each model's performance on the testing set using metrics such as accuracy score, recall score, precision score, f1_score, and support. Accuracy score measures the fraction of correct predictions on the test set, while other metrics offer insights into different aspects of the dataset.

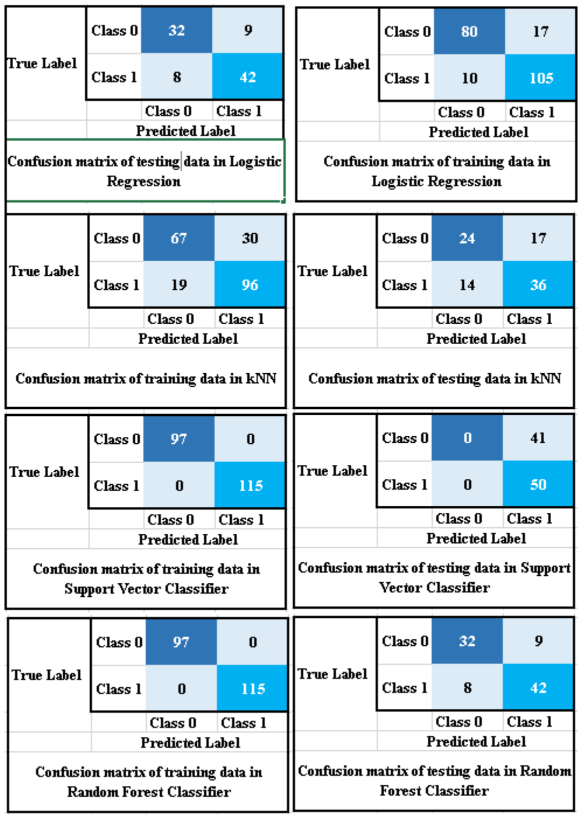

Fig 3. Confusion Matrix of various classifiers

It evaluates every model's performance on the training set to perceive over fitting. The Logistic Regression, K-Neighbours Classifier, SVM, and RF exhibit the maximum testing accuracy between 54.95 to 81.32 %.

A confusion matrix is crucial in evaluation and assess accuracy of classification model by presenting results in a two-by-two matrix. It includes counts of false positives and negatives, true positives and negatives, offering insight into model performance disparities to predict positive and negative instances.

The comprehensive analysis of the dataset based on Cleveland Heart Disease through AI, includes metrics like precision, recall, f1, and support, along with the confusion matrix. It provides a thorough assessment of model performance and serves as a valuable starting point for further research on MI.

The confusion matrix in Figure 3 is indicated for various classifiers including different techniques of testing and training.

5.1Comparative Analysis

In multiple AI models, the dataset at Cleveland Heart Disease is being utilized to judge MI. It includes different models like Logistic Regression, K-NN, SVC, and Random Forest Classifiers. One of the primary techniques used in the method is pre-processing the data by eliminating mislaid values, encoding categorical variables, as well as feature scaling. It is essential to prepare the data for AI models as it helps in eradicating discrepancies as well as errors in the dataset.

To assess the effectiveness, the several metrics such as precision, accuracy are used with recall, f1_score, and support. The accuracy determines the proportion of accurate predictions of the model on the test set, while the other metrics gauge the models' performance on numerous aspects of the dataset.

The quality of each model on the training set determine over fitting. It is essential to analyse since an excessive fitting of the data can lead to the below-average test performance of the model regardless of the satisfactory training results.

The metrics used for classification offer an extensive assessment of the models functioning. For instance, precision determines the percentage of correct positive predictions made by the proposed model from the entire amount of positive forecasts, while recall gauges the part of accurate positive forecasting's made by the model from the entire number of existing positive results.

6. Conclusion and future Scope

6.1Conclusion

Predicting MI is a serious and complicated issue. Because of the complexity of the disease and the potential severity of its consequences, it is a critically important to accurately predicting and preventing MI. Accurately predicting who is at risk for MI requires the analysis of a wide range of data points, including medical history, lifestyle factors, and genetic predisposition. It can be difficult to diagnose and treat, and risk factors for developing MI are complex and multifaceted and manual methods for assessing risk factors are difficult and unreliable.

Using the system explained in this proposed research paper which applies AI techniques and various algorithms to forecast the probability of cardiac disease based on the information. One limitation of this approach is that it focuses primarily on classifying techniques and algorithms, and does not address other aspects of MI diagnosis and treatment. Despite the discussed limitation, the system can be widely used and accessible, even by non-medical professionals. It can also help to reduce the workload and time complexity of doctors. It is necessary to in this area, and continue research and innovation to enhance the MI prediction accuracy and outcomes.

6.2Future scope

A study has potential for further innovation as well as enhancing the field of MI prediction. While the current algorithms used in the study are efficient, other AI techniques and potential algorithms could also be applied to enhance the precision of MI prediction like deep learning, decision trees, or other advanced models.

Along with developing new techniques and algorithms, there is also scope for improving efficiency with more and more data collection and analysis. Finding ways to rapidly process this data with higher accuracy make more timely and informed predictions.

Furthermore, there is still much to be explored and discovered in the area of prediction of MI. The continued investigation and innovation revise the exactness and efficiency of the area of healthcare.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

[1] B. Ziaeian and G. C. Fonarow, "Epidemiology and aetiology of heart failure," *Nat. Rev. Cardiol.*, vol. 13, no. 6, pp. 368–378, 2016, doi: 10.1038/nrcardio.2016.25.

[2] T. Vos *et al.*, "Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015," *Lancet*, vol. 388, no. 10053, pp. 1545–1602, 2016, doi: 10.1016/S0140-6736(16)31678-6.

[3] E. J. Benjamin *et al.*, *Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association*, vol. 139, no. 10. 2019. doi: 10.1161/CIR.0000000000000659.

[4] V. L. Roger *et al.*, "Heart disease and stroke statistics-2012 update: A report from the American heart association," *Circulation*, vol. 125, no. 1, 2012, doi: 10.1161/CIR.0b013e31823ac046.

[5] C. A. Lawson *et al.*, "Risk factors for heart failure: 20-year population-based trends by sex, socioeconomic status, and ethnicity," *Circ. Hear. Fail.*, no. February, 2020, doi: 10.1161/CIRCHEARTFAILURE.119.006472.

[6] S. M. Artham, C. J. Lavie, R. V. Milani, and H. O. Ventura, "Obesity and hypertension, heart failure, and coronary heart disease - Risk factor, paradox, and recommendations for weight loss," *Ochsner J.*, vol. 9, no. 3, pp. 124–132, 2009.

[7] F. D. Fuchs and P. K. Whelton, "High Blood

Pressure and Cardiovascular Disease," *Hypertension*, no. Cvd, pp. 285–292, 2020, doi: 10.1161/HYPERTENSIONAHA.119.14240.

[8] A. Lala and A. S. Desai, "The role of coronary artery disease in heart failure," *Heart Fail. Clin.*, vol. 10, no. 2, pp. 353–365, 2014, doi: 10.1016/j.hfc.2013.10.002.

[9] M. E. Piché, A. Tchernof, and J. P. Després, "Obesity Phenotypes, Diabetes, and Cardiovascular Diseases," *Circ. Res.*, pp. 1477–1500, 2020, doi: 10.1161/CIRCRESAHA.120.316101.

[10] E. M. McNally and L. Mestroni, "Dilated cardiomyopathy: Genetic determinants and mechanisms," *Circ. Res.*, vol. 121, no. 7, pp. 731–748, 2017, doi: 10.1161/CIRCRESAHA.116.309396.

[11] K. Klingel, M. Sauter, C. T. Bock, G. Szalay, J. J. Schnorr, and R. Kandolf, "Molecular pathology of inflammatory cardiomyopathy," *Med. Microbiol. Immunol.*, vol. 193, no. 2–3, pp. 101–107, 2004, doi: 10.1007/s00430-003-0190-1.

[12] S. Pankuweit and K. Klingel, "Viral myocarditis: From experimental models to molecular diagnosis in patients," *Heart Fail. Rev.*, vol. 18, no. 6, pp. 683–702, 2013, doi: 10.1007/s10741-012-9357-4.

[13] M. R. Piano, "Alcohol's Effects on the Cardiovascular System," *Alcohol Res.*, vol. 38, no. 2, pp. 219–241, 2017.

[14] P. Ponikowski *et al.*, "2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure," *Eur. Heart J.*, vol. 37, no. 27, pp. 2129-2200m, 2016, doi: 10.1093/eurheartj/ehw128.

[15] Y. Xia and J. Wang, "A One-Layer Recurrent Neural Network for Support Vector Machine Learning," *IEEE Trans. Syst. Man, Cybern. Part B Cybern.*, vol. 34, no. 2, pp. 1261–1269, 2004, doi: 10.1109/TSMCB.2003.822955.

[16] M. Gandhi and S. N. Singh, "Predictions in heart disease using techniques of data mining," *2015 1st Int. Conf. Futur. Trends Comput. Anal. Knowl. Manag. ABLAZE 2015*, pp. 520–525, 2015, doi: 10.1109/ABLAZE.2015.7154917.

[17] S. Harikrishnan *et al.*, "Five-year mortality and readmission rates in patients with heart failure in India: Results from the Trivandrum heart failure registry," *Int. J. Cardiol.*, vol. 326, pp. 139–143, 2021, doi: 10.1016/j.ijcard.2020.10.012.

[18] J. Nahar, T. Imam, K. S. Tickle, and Y. P. P. Chen, "Association rule mining to detect factors which contribute to heart disease in males and females," *Expert Syst. Appl.*, vol. 40, no. 4, pp. 1086–1093, 2013, doi: 10.1016/j.eswa.2012.08.028.

[19] L. Ali *et al.*, "An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure," *IEEE Access*, vol. 7, pp. 54007–54014, 2019, doi: 10.1109/ACCESS.2019.2909969.

[20] Y. Udovychenko, A. Popov, and I. Chaikovsky, "K-NN binary classification of heart failures using myocardial current density distribution maps," *2015 Signal Process. Symp. SPSympo 2015*, no. Cddm, 2015, doi: 10.1109/SPS.2015.7168283.

[21] S. W. P. M. D. R. & M. D. M. D. Janosi Andras, "Heart Disease." 1988.

[22] R. Santosa, A. C.-R. World, and undefined 2017, "Logistic Regression Model for Predicting First Semester Students GPA category Based on High School Academic Achievement," *researchgate.net*, 2017, doi: 10.18843/rwjasc/v8i2/01.

[23] I. R. A. Hamid, N. S. Khalid, N. A. Abdullah, N. H. A. Rahman, and C. C. Wen, "Android Malware Classification Using K-Means Clustering Algorithm," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 226, no. 1, 2017, doi: 10.1088/1757-899X/226/1/012105.

[24] Q. Feng, J. Liu, and J. Gong, "Urban flood mapping based on unmanned aerial vehicle remote sensing and random forest classifier-A case of yuyao, China," *Water (Switzerland)*, vol. 7, no. 4, pp. 1437–1455, 2015, doi: 10.3390/w7041437.