Minimum Skewness based Myocardial Infarction Detection Model using Classification Algorithms

Original Scientific Paper

Satish Kumar Tewalker*
Department of IT and Computer Application
Dr. C. V. Raman University
Kota Bilaspur,India
satishkumarphd88@gmail.com

Ragini Shukla
Department of IT and Computer Application
Dr. C. V. Raman University
Kota Bilaspur,India
raginishukla008@gmail.com

*Corresponding author

Abstract *– Myocardial infarction is one of the most dangerous public health issues in the world. The accurate prediction of myocardial infarction disease aids in disease diagnosis and biological analysis of the patient's health. The classification algorithms are one of the solutions that predict accurate diseases based on the symptoms (attributes) in patients' details. The ability to predict accurately reduces the risk of causality and decision-making time. This study proposed the Minimum Skewness-Based Myocardial Infarction Detection Model (MSMIDM) with the help of a statistical and feature selection-based approach. Minimum skewness is a feature selection statistical approach that selects essential attributes of a dataset. The MSMIDM provides accurate results with the highest accuracy among the six classification algorithms. The experimental analysis makes use of the most widely used Cleveland dataset for myocardial infarction detection. The experimental results are analyzed through a confusion metric, statistical, and partitional validation approach. The proposed model obtains an accuracy of 90%, 87.037%, 87.037%, 83.238%, 81.481%, and 85.556% with respect to Random Forest, K-Nearest Neighbor, Support Vector Machine, Naive Bayes, Decision Tree and Neural Network classification algorithms. According to the experimental analysis, this study suggests the MSMIDM-based Random Forest algorithm is excellent for myocardial infarction disease detection.*

Keywords*:* *Heart Diseases, Statistical Classification, Feature Selection, Skewness Classification, Myocardial Infarction Detection Model*

Received: April 20, 2024; Received in revised form: June 17, 2024; Accepted: July 2, 2024

1. INTRODUCTION

Heart disease is the most dangerous disease nowadays that impacts numerous individuals in their middle or late years and leads to deadly consequences for many causes. According to WHO figures, cardiac disorders account for 24% of all non-communicable disease fatalities in India [1]. Heart disorders account for one-third of all global fatalities [2] [3]. Around 17 million individuals die each year as a result of cardiovascular disease (CVD), with the ailment being especially prevalent in Asia [4]. Tobacco use, cholesterol, high blood pressure, family history, poor diet, alcohol consumption, eating habits, physical inactivity, diabetes and obesity are all known heart disease risk aspects [5].

Myocardial infarction (MI) is the most widespread heart disease and it is the most prevalent cause of death and morbidity, as well as a significant cost of care because it is a major cause of mortality and disability all around the world [6]. Myocardial Infarction generally recognized as a heart attack, which is one of the most dangerous and deadly cardiovascular diseases [7]. Existing research indicates a relationship between the formation of MI disease and a blockage in the coronary artery. The first clinical accounts arose in the early twentieth century as clinical symptoms were connected to the coronary artery. Denmark et al. (2019) [6] depicted the five principal varieties of MI. The first is a coronary incident. The second cause of MI is caused by an oxygen supply. The third case is a fatal cardiac death. Percutaneous coronary intervention is the fourth cause of MI, and the last significant type is connected to coronary artery bypass grafting.

Manually determination of risk factors for myocardial infarction and other diseases is a challenging task [8]. Machine learning techniques help extract hidden risk factors for myocardial infarction heart diseases with the help of existing data and high accuracy. Machine learning is used to extract hidden information, facts, patterns, and knowledge related to diseases from a variety of disease-related data sources. Machine learning determines these sorts of patterns by employing intelligible functions and procedures. In general, machine learning algorithms are classified as either predictive or descriptive. Predictive learning employs inference to drive conclusions from data sources. Descriptive learning is the differentiation of broad properties in any data storage [9].

Classification is one of the most essential and effective approaches in predictive learning, which is known as supervised machine learning. Classification is essential for dealing with massive amounts of data, and it is used for data analysis to predict class labels in the form of categorical labels [10]. The classifier employed two steps for predicting the class label. The first step is learning, where a predicated class is produced from a dataset of known classes. The second step predicts the class labels based on the dataset. The second step is an important task that is used by the constructed classifiers [11]. There are several data mining classification techniques including Rule-Based Classifier, Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Naive Bayes (NB) and Neural Network (NN).

Myocardial infarction is a leading cause of mortality and disability around the world. The prediction of myocardial infarction disease is beneficial for reducing the death ratio, and machine learning assists in identifying the voluntary history of many diseases, including myocardial infarction diseases of the heart. Researchers employed well-known biostatistical approaches to predict the relationship between MI and associated risk factors in patients. Biostatistical approaches are inappropriate for identifying correct data and risk indicators. There is a global need to use more appropriate machine learning algorithms with statistical techniques in the medical field for disease prediction. The objective of this research is to predict MI among chronic diseases through a classification approach to extract knowledge from data with the help of statistical techniques.

The rest of the paper is structured as follows: Section 2 reviews the various classification models for myocardial infarction detection. Section 3 details the objective function of classification and proposes the classification model through a feature selection statistical approach. Section 4 implements the proposed classification model and compares its performance with various classical classification models and algorithms. Section 5 concludes the study and describes further work.

2. RELATED WORK

The classification model is essential in predicting myocardial infarction heart disease using various statistical approaches, feature selection algorithms, optimization metrics, etc. This section examined the literature related to several investigations and classification models for myocardial infarction heart disease prediction.

Parthiban and Srivatsa [12] select the attributes with the help of correlation and predict myocardial infarctions through SVM and NB classification methods.

Sun et al. [13] constructed an automatic detection system for myocardial infarction in patients utilizing supervised machine learning techniques known as multiple instance learning (MIL). The experimental results of the proposed MIL algorithm are assessed using the classical MIL method with the help of the ECG and PTB diagnostic databases.

Safdar et al. [14] investigated seven classification and association rules learning algorithms to produce an accurate risk analysis prediction model in myocardial infarction. The experiment findings are compared utilizing validation models for accuracy, sensitivity, specificity and precision. The CHAID decision tree model outperforms other classification techniques by obtaining 93.4% precision.

Seenivasagam et al. [7] suggested a computer-aided detection method that assesses a supervised machine learning classifier to assess the risk level of myocardial infarction. The experimental investigation demonstrated that the PSONN classifier beat other algorithms regarding risk prediction, while the neural network classifier exceeded other algorithms regarding dataset training accuracy.

Otoom et al. [15] applied hold-out and cross-validation tests and used NB, SVM, and DT classification algorithms to predict myocardial infarction heart diseases. Vembandasamy et al. [16] employ the various computational parameters of the NB method to predict heart diseases.

Daraei et al. [17] developed the model for MI detection using the cost-sensitive J48 approach. It employed a hybrid feature selection approach for feature selection, whereas the meta cost classifier to prediction. The experimental research revealed that combining hybrid feature selection and metacost classifier enhances sensitivity and obtains a greater propensity for MI prediction compared to classical classification.

Verma et al. [18] use a hybrid method that includes correlation-based feature selection, K-means clustering, particle swarm optimization, multinomial logistic regression, multi-layer perceptrons and fuzzy unordered rule induction algorithm to predict myocardial infarction. Chadha and Mayank [19] extract disease patterns from vital data and utilize NN, DT, and NB classification algorithms to extract disease patterns.

Dwivedi et al. [20] established a classification framework for detecting heart disease utilizing tenfold cross-validation and machine-learning algorithms. The experimental analysis shows that the logistic regression classification algorithm achieves excellent levels of classification accuracy (85%), sensitivity (89%) and specificity (81%) as compared to other classification algorithms.

Mokeddem et al. [21] developed a clinical decision support system (CDSS) to predict MI and provide diagnosis for Coronary Artery Diseases (CAD). The proposed diagnosis model used RF, C5.0 decision tree and fuzzy classification algorithms. The RF method is used to extract features based on rank, the C5.0 decision tree is used to generate crisp rules, and the fuzzy algorithm is used to forecast disease and diagnosis. The CDSS model has the capability to handle missing and noisy data. The experimental research revealed that the CDSS optimized processing time and provided a classification accuracy of up to 90.50%.

Ed-daoudy and Maalmi [22] uses the well-known SVM, DT, RF and Logistic Regression classification algorithms and implement inside the Apache Spark for big data processing. The proposed model improved accuracy and computing time while suggesting a logical resolution for detecting myocardial infarction disease. Anitha and Sridevi [23] used SVM, KNN, and NB classification algorithms to improve the accuracy of the heart disease detection framework. The experiment results indicate that NB is excellently suited for detecting myocardial infarction disease.

Christalin Latha et al. [1] enhance the accuracy of classification algorithms to detect heart disease using ensemble classifiers. It uses ensemble techniques such as stacking, bagging, majority voting, and boosting. The bagging method increases accuracy by 6.92%. The boosting method increases accuracy to 5.94. The majority voting method increases accuracy by 7.26% and the stacking method increases accuracy to 6.93%.

Garate-Escamila et al. [24] suggested a dimensionality reduction approach for detecting risk characteristics of heart disease. The suggested dimensionality reduction approach is the CHI-PCA algorithm. It employs PCA for feature extraction and CHI for feature selection. The experimental findings demonstrated that CHI-PCA with the random forest classification algorithm achieves the maximum accuracy in various data sets.

Shah et al. [25] assessed the effectiveness of several supervised machine learning methods to detect the probability of heart disease in patients, such as myocardial infarction. The experimental results demonstrated that the KNN classifier outperformed the NB, DT, and RF algorithms in accuracy.

Mandair et al. [26] developed a machine-learning model to predict MI utilizing random sampling and a deep neural network. The experimental study employed harmonized Electronic Health Record data and seven layers of one hundred neurons per layer. According to the experimental results, the suggested model has a specificity of 73.3%, a sensitivity of 82%, a recall of 82%, and an accuracy of 05%.

Ibrahim et al. [27] developed the active MI prediction framework using a synthetic minority oversampling technique (SMOTE), recurrent neural network (RNN), decision-tree, convolutional neural network (CNN) and XGBoost classification methods. The suggested model achieves 89.9%, 84.6%, and 97.5% accuracy.

Reddy et al. [28] developed a forecasting system for myocardial infarction utilizing hybrid machine learning methods. The suggested approach outperforms the KNN, DT, NN, NB, RF and SVM classifiers in terms of accuracy. The experiment findings reveal that the random forest method achieves a superior accuracy compared to other classical algorithms.

Patro et al. [29] designed a heart disease prediction framework utilizing risk attributes and Salp Swarm Optimized Neural Network (SSA-NN), NB, KNN, and Bayesian Optimized Support Vector Machine (BO-SVM) classification algorithms. The experimental results revealed that the proposed framework optimized the classification algorithm and provided an excellent and effective healthcare monitoring system.

Kondababu et al. [30] integrated the properties of RF and linear method and constructed a hybrid random forest-linear approach to predict myocardial infarction disease. Liu et al. [31] presented a risk prediction model for actively monitoring patients with myocardial infarction disease. Top-layer and bottom-layer algorithms are used in the suggested risk prediction model. The top layer contains the recursive feature elimination technique, while the bottom layer has the gradient boosting decision tree (GBDT), RF, SVM, and logistic regression algorithms.

Nagavelli et al. [32] improved the machine-learning model that detects heart disease. The improved model employs the Naïve Bayes classifier with weighted models, SVM classifier with XGBoost algorithm. The experimental investigation demonstrates that the single XGBoost algorithm obtains excellence in permanence, whereas the Naïve Bayes classifier obtains the lowest accuracy when compared to other algorithms.

Sadiyamole et al. [33] recommended combining the advantages of genetic algorithms with ensemble deep learning to predict heart disease. The genetic algorithm retrieves relevant attributes from the dataset, while the deep learning technique predicts the presence of heart disease. The proposed method achieves excellent accuracy as compared to the classical deep learning approach.

Ahmad et al. [34] used the jellyfish method to optimize the machine-learning model. The jellyfish method extracts the dataset's essential features and uses machine learning to predict heart illnesses. The experiment analysis demonstrates that the SVM classifier outperforms other machine learning techniques. The suggested model addresses model complexity and overfitting-related issues.

Bhatt et al. [35] employ a clustering and classification classifier to accurately predict cardiovascular diseases. The suggested model first employs K-mode clustering to classify the data before beginning the classification procedure. In classification, the data was divided into 80:20 ratios for building models, and a number of classifiers were utilized to predict heart disease. The experimental results demonstrate that K-mode clustering and multilayer perceptrons are more accurate.

Subramani et al. [36] implemented a cardiovascular disease detection model in a real-time environment using the Internet of Things. The proposed model first collects the data from IoT devices and then preprocesses the data according to the classification model. The classification model observed the data and automatically decided the training and testing classifiers. The experimental analysis shows that the deep learning model achieves excellent performance as compared to other classifiers.

Al Alshaikh et al. [37] improved the cardiovascular heart disease prediction model by combining different feature extraction-related algorithms and machine learning classifiers. The improved model initially employs feature-related algorithms. The proposed model's second stage includes classification tasks performed by a multilayer deep convolutional neural network. The multilayer deep convolutional neural network outperforms another classifier using the adaptive elephant herd optimization approach.

3. PROPOSED ALGORITHM

This section states the classification objective and describes the Minimum Skewness Myocardial Infarction Detection Model (MSMIDM) using a different classification algorithm.

3.1. CLASSIFICATION OBJECTIVE FUNCTION

The preprocessed MDS myocardial infarction diseases dataset consists of $X=\{x_1, x_2, x_3 \ldots x_N\}$ data points with $X_d=\{x_{1d}, x_{2d}, x_{3d} \ldots x_{Nd}\}$ attributes of myocardial infarction diseases. The dimensions' attributes defined the disease's behavior, symptoms, and nature of the sample data. The classification algorithms divide the MDS dataset into various classes based on the symptoms of diseases. The predicted class describes the severity of myocardial infarction diseases and suggests a diagnosis through a consultation with a physician. The goal of classification is to maximize predicted class accuracy while minimizing diagnosis and analysis time for myocardial infarction disease patients.

Suppose $C=\{c_1, c_2, c_3 \ldots c_k\}$ is the predicted class of the classification and $P=\{p_1, p_2, p_3 \ldots p_{k'}\}$ is the predefined class of the myocardial infarction diseases dataset. The K and K' represented the number of classes related to diseases. The number of classes of classification results and predefined class is equal and their intersection is always null. The $N_{i,j}$ is the number of correctly classified data that belongs to both predicted and predefined classes. The N_i is the number of data points of the predicted class of the classification. The N_j is the number of data points that belong to only a pre-defined class. Thus, the objective of the classification is to maximize the $N_{i,j}$ number of data points to obtain better accuracy. The formulation of the objective function is shown in Equation (1) to (3), respectively {Formatting Citation}.

$$Max = \sum_{i=1}^{k} max_{j=1}^{K'} N_{i,j} \quad (1)$$

$$s.t. \quad min\, N_j \; and \; N_i \quad (2)$$

$$N_j \cap N_i = \emptyset \quad (3)$$

3.2. CLASSIFICATION ALGORITHM

Machine learning algorithms utilize a diverse range of statistical, mathematical, probabilistic, and optimization-related methods to learn from previous experience and detect meaningful patterns, knowledge, and trends in large, unstructured, and complex disease datasets. This study uses Random Forest (RF), Support Vector Machine (SVM), Naive Bayes(NB), K-Nearest Neighbor (KNN), Decision Tree (DT), Neural Network (NN) machine learning algorithms for myocardial infarction disease detection [20, 25, 38-41].

3.3. PROPOSED MYOCARDIAL INFARCTION DETECTION MODEL

This section describes MSMIDM in three stages. The first phase of MSMIDM is to select the most significant adverse skewness attributes by utilizing the mean and standard deviation of the myocardial infarction disease datasets. The selected attributes avoid outlining and noise in the dataset during the classification. The second phase used a different classification algorithm to predict myocardial infarction disease with the help of selected negative skewness attributes.

The third phase selects the results with the highest accuracy across all classification algorithms. Thus, the first phase identifies the probable attributes of the dataset, the second phase assesses the accuracy of all selected classification algorithms, and the last phase generates correct results for myocardial infarction disease.

3.3.1 Skewness

Skewness is a third-moment statistical approach that analyzes the symmetry and accuracy of a dataset utilizing the normal distribution's probability distribution. The skewness of the data is classified as positive or negative depending on its skewed curve. Inside a positive (maximum) skewness, the mean value is more significant than the mode value. Therefore, the data is skewed towards the right side of the curve. The mode value exceeds the mean within the negative (minimal) skewness.

Therefore, the data is skewed towards the left side of the curve. The normal distribution selects those data points that are shifted to higher side. If most of the data is moved inside a positive curve, it indicates that data is acceptable for classification. If most of the data is moved into a negative curve, it indicates that data is used for classification [42, 43]. Here, the most negative data are selected for classification that reason this study divided the data into 80:20 ratios. If we select the 80 % negative data during the classification, then the data is divided into 70:30 ratios. Thus, the data division ratio is dependent on the skewness value and it is validated by the K-fold cross-validation.

The Fisher-Pearson coefficient formulation of skewness was employed in this investigation. The equation (4) depicts the formulation of skewness.

$$S = \frac{\sum(x_i - \bar{x})^3/n}{\sigma^3} = \frac{3\,(\mu - m)}{\sigma} \quad (4)$$

Where x_i is the data point of the selected attributes of the dataset, $\bar{x}$ and μ is the mean of the selected attribute of the dataset, n is the total number of data points inside a selected attribute of the dataset, m is the median of the selected attribute of the dataset, σ_3 is the third moment of standard deviation for the selected attributes, σ is the standard deviation of a specific attribute of the dataset.

3.3.2. Algorithm Description

The Minimum Skewness-based Myocardial Infarction Detection Model (MSMIDM) is described in this subsection through skewness, classification algorithms, and accuracy measurement for myocardial infarction detection. The algorithm and flow chart of the proposed MSMIDM algorithm are outlined in Algorithm 1 and Fig. 1, which find excellent accuracy in the myocardial infarction disease prediction.

4. EXPERIMENTAL ANALYSIS

This section implements the proposed model for myocardial infarction heart disease prediction with the help of an experimental environment, myocardial infarction disease dataset, characteristics of the dataset, evaluation criteria and results. The experiment analysis of this study examined the accuracy of the proposed model and state-of-the-art models.

The implementation code of MSMIDM is written using Python 3.5.3, a programming language inside the Jupyter Notebook. The computational environment of the system is configured with a 320 GB hard disk, 4 GB main memory, Intel I3 processor, CPU M350@2.27 GHz, and Windows 7 operating system.

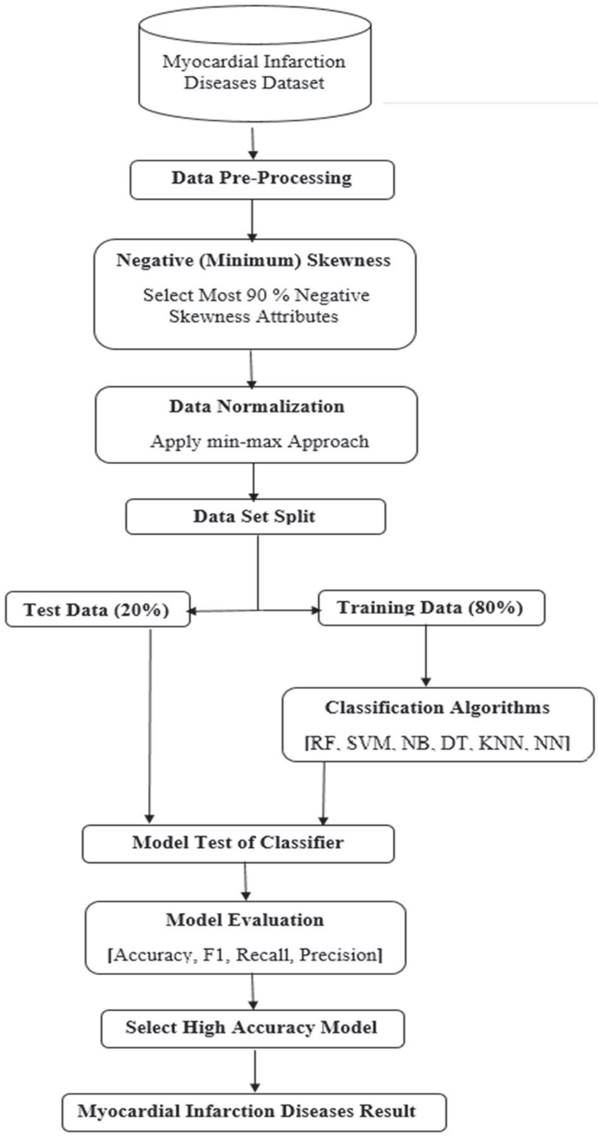

Figure 1. Flow Chart of the Proposed Minimum Skewness-based Myocardial Infarction Detection Model (MSMIDM)

4.1. EXPERIMENT DATASET AND ALGORITHM

This study utilizes the Cleveland Myocardial Infarction Diseases dataset from the UCI Machine Learning Repository. The used dataset consists of 300 samples with 14 dimensions, where 13 dimensions define predictors and 1 dimension describes the actual class. The details of this dataset are described in various studies, which is described in Table 1. The proposed Myocardial Infarction Detection Model (MSMIDM) was compared to various state-of-the-art models based on classification objectives. The description of the selected classification model is described in Table 1 references.

Algorithm 1: Minimum Skewness based Myocardial Infarction Detection Model (MSMIDM)

Input:

1. *MDS*={x_1, x_2, …x_n} *MI Diseases Dataset with d dimensions*.

2. *CA = Set of Classification Algorithms [RF, SVM, NB, DT, KNN, NN]*

Output:

1. *Myocardial Infarction Diseases results basis of CA.*

Method:

1. *Perform data preprocessing tasks based on missing values, and so on.*
2. *Extract most of the 90 % negative skewness attributes for classification through Equation (1).*
3. *Perform the min-max normalization approach for data normalization.*
4. *Splitting the data into test and training data based on 80:20 ratios as per selection of negative skewness attributes.*
5. *Incorporate the CA classifier set into the training data.*
6. *Test the model based on the classifier results.*
7. *Evaluating the model based on classification validation approaches.*
8. *Select a highly accurate model for myocardial infarction disease analysis.*

Table 1. Description of Selected Algorithms for Accuracy Comparison

ID	Algorithms/Authors	References
A01	(Parthiban and Srivatsa, 2012)	[13]
A02	(Otoom et al., 2015)	[16]
A03	(Vembandasamy et al.,2015)	[17]
A04	(Verma et al.,2015)	[20]
A05	(Chadha and Mayank, 2016)	[21]
A06	(Seenivasagam and Chitra, 2016)	[7]
A07	(Dwivedi, 2018)	[23]
A08	(Ed-daoudy and Maalmi, 2019)	[25]
A09	(Anitha and Sridevi, 2019)	[26]
A10	Bagging (Latha and Jeeva, 2019)	[1]
A11	Boosting (Latha and Jeeva, 2019)	[1]
A12	(Shah et al., 2020)	[28]
A13	(Patro et al., 2021)	[33]
A14	(Reddya and G,2021)	[32]
A15	(Kondababu et al.,2021)	[34]

4.2. PERFORMANCE MEASURES

The performance metrics assess the perfection of algorithms and models. This study utilized accuracy, F1-measure, recall, precision, Jaccard score, Mutual Info Score (MIS), Adjusted Rand Score (ARS), V Measure Score (VMS), Mean Squared Error (MSE) and R2 Score (R2) performance metrics to evaluate the proposed model. The formulation of all employed measurements is described in respectively [20], [25], [38-41], [44], and [45].

4.3. RESULTS AND DISCUSSION

The performance of numerous classification methods is compared in Tables 2-4. Table 2 depicts the confusion matrix-based accuracy, F1 measure, recall and precision measurement of various classification algorithms employing the proposed model. The RF algorithm obtains the maximum accuracy of 90%, the F1 measure of 88.1%, the recall of 83.333%, and the precision of 78.72% in Table 2 compared to the state-of-the-art algorithms.

Table 3 compares the confusion matrix and statistically based ARS and VMS measurements for the proposed model. The RF algorithm achieves the highest ARS of 63.3% and VMS of 53.97% in Table 3 as compared to the state-of-the-art algorithm.

Table 2. Classification results of the MSMIDM-based Machine Learning(ML) approach using confusion matrix component

ML Algorithms	Accuracy (%)	F1 (%)	Recall (%)	Precision (%)
RF	90	88.101	83.333	78.727
SVM	87.037	84.444	79.167	73.077
NB	87.037	84.444	79.167	73.077
DT	83.238	79.475	79.167	65.977
KNN	81.481	77.273	70.833	62.963
NN	85.556	82.976	79.167	70.912

Table 3. Classification results of the MSMIDM-based Machine Learning(ML) approach using confusion matrix and statistical component

ML Algorithms	ARS (%)	VMS (%)
RF	63.305	53.971
SVM	54.027	44.973
NB	54.027	44.973
DT	39.538	31.437
KNN	38.524	31.241
NN	49.664	40.441

Table 4 compares the Jaccard score, MIS, MSE, and R2 classification results of the RF, SVM, NB, DT, KNN, and NN classification algorithms applying a partitional validation method to the proposed model. In this case, the RF method obtains the most robust Jaccard score of 78.73%, MIS of 36.36%, and R2 of 59.5% compared to the state-of-the-art approach. The SVM and NB achieve outstanding classification results within an MSE validation index. The MSE for RF results demonstrates that randomization influences classification outcomes over non-randomization-based classification techniques.

Tables 5-10 examine the comparative accuracy of the proposed model to that of state-of-the-art models. Table 5 reveals that the proposed MSMIDM-based RF algorithm delivers more robust accuracy results than the existing work on RF-based models. The MSMIDM-based RF classifier obtained 90% as compared to the state-of-the-art model. The MSMIDM improved accuracy by 3.16% when compared to the excellent A12 algorithm. The MSMIDM improves worst-case RF accuracy by 11.12% compared to the A11 model.

Table 4. Classification results of the MSMIDM-based Machine Learning(ML) approach based on partitional validation approach

ML Algorithms	Jaccard Score (%)	MIS (%)	MSE (%)	R2 (%)
RF	78.727	36.358	13.334	59.5
SVM	73.077	30.474	12.963	47.5
NB	73.077	30.474	12.963	47.5
DT	65.977	21.57	18.148	26.5
KNN	62.963	21.027	18.519	25
NN	70.912	27.523	14.444	41.5

Table 5. Accuracy assessment of the proposed MSMIDM and conventional models for RF algorithm

Algorithms	Accuracy (%)
A08	87.5
A10	80.53
A11	78.88
A12	86.84
A15	86.1
MSMIDM	**90**

Table 6 illustrates that the proposed MSMIDM-based SVM algorithm produces excellent classification accuracy results compared to other SVM-based models. The MSMIDM-based SVM classifier outperformed the most recently developed model by 87.037%. The MSMIDM improved the accuracy by 0.937% when compared to the superior A15 algorithm. The MSMIDM resolves the worst-case accuracy of SVM using the A14 model and improves the accuracy by 12.037% over the A14 model.

Table 6. Accuracy assessment of the proposed MSMIDM and conventional models for SVM algorithm

Algorithms	Accuracy (%)
A06	82.02
A02	84.50
A07	82
A08	85.82
A09	77.7
A13	80
A14	75
A15	86.1
MSMIDM	**87.037**

Table 7 shows that the proposed MSMIDM-based NB algorithm achieves reasonable accuracy when compared to earlier NB-based models. The MSMIDM-based NB classifier outperformed the state-of-the-art model by 87.037%. The MSMIDM improved the accuracy by 0.037% against the outstanding A13 algorithm. The MSMIDM improves the worst-case accuracy of NB by 13.037% as compared to the A01 model.

Table 8 shows that the proposed MSMIDM-based DT algorithm achieves superior classification results than former DT-based models in terms of accuracy validation. The MSMIDM-based DT classifier outperformed the state-of-the-art model by 83.238%, while MSMIDM improved accuracy by 0.438% over the effective A08 strategy. The MSMIDM resolves the worst-case accuracy of DT using the A11 model and improves accuracy by 7.338% with the A11 model.

Table 7. Accuracy assessment of the proposed MSMIDM and conventional models for NB algorithm

Algorithms	Accuracy (%)
A01	74
A02	84.50
A03	86.42
A05	85.86
A07	83
A09	86.6%
A10	84.16
A11	84.16
A13	86.7
A14	78
A15	75.8
MSMIDM	**87.037%**

Table 8. Accuracy assessment of the proposed MSMIDM and conventional models for DT algorithm

Algorithms	Accuracy (%)
A04	80.6
A07	77
A08	82.8
A10	79.87
A11	75.9
A12	80.263
MSMIDM	**83.238**

Table 9 shows that the proposed MSMIDM-based KNN algorithm has the highest accuracy outcomes when compared to other KNN-based models. The MSMIDM-based KNN classifier outperformed the state-of-the-art model by 81.481%. The MSMIDM boosted accuracy by 1.484% when compared to the most effective A07 and A13 algorithms. The MSMIDM resolves the worst-case accuracy of KNN using the A14 model and improves the accuracy by 11.481% over the A14 model.

Table 9. Accuracy assessment of the proposed MSMIDM and conventional models for KNN algorithm

Algorithms	Accuracy (%)
A07	80
A09	76.67
A13	80
A14	70
MSMIDM	**81.481**

Table 10 shows that the proposed MSMIDM-based NN algorithm produces better classification results than other NN-based models regarding accuracy validation. The MSMIDM-based NN classifier outperformed the most advanced model by 85.556%. The MSMIDM improved accuracy by 1.553% when compared to the A07 algorithm. The MSMIDM improves the worst-case accuracy of NN by 3.016% when compared to the A11 model.

Table 10. Accuracy assessment of the proposed MSMIDM and conventional models for NN algorithm

Algorithms	Accuracy (%)
A07	84
A10	81.52
A11	79.54
A13	80
A14	82
MSMIDM	**85.556**

Figs. 2–7 depict a comparison of conventional and proposed model-based algorithms utilizing accuracy, F1-measure, recall, precision, Jaccard score, MIS, ARS, VMS, MSE, and R2 classification metrics.

Fig. 2 illustrates that the MSMIDM-based RF algorithm maximizes the confusion, statistical, and partitional-based validation metrics compared to the standard RF method. The MSMIDM-based RF classifier improves the 1.67 % accuracy, 1.69 % F1-measure, 2.64 % precision, 2.64% Jaccard score, 3.46% MIS, 5.24% ARS, 5.74% VMS, and 6.75% R2 validation metrics as compared to the classical RF classifier.

Fig. 3 shows that the proposed MSMIDM-based SVM method outperforms classical SVM regarding all considered validation approaches. The MSMIDM-based SVM enhanced the 5.56 % accuracy, 5.28 % F1-measure, 0.0 % recall, 7.56 % precision, 7.56 % Jaccard score, 9.55 % MIS, 15.52 % ARS, 14.52 % VMS, and 22.5 % R2 assessment as compared to the classical SVM classifier.

Fig. 4 depicts that the MSMIDM-based NB algorithm outperforms the classical NB approach based on classification results. The MSMIDM-based NB outperformed the standard NB classifier in terms of 7.41% accuracy, 8.89% F1-measure, 8.33% recall, 12.36% precision, 12.36% Jaccard score, 12.29% MIS, 20.12% ARS, 18.14% VMS, and 30.0% R2 authentication metric.

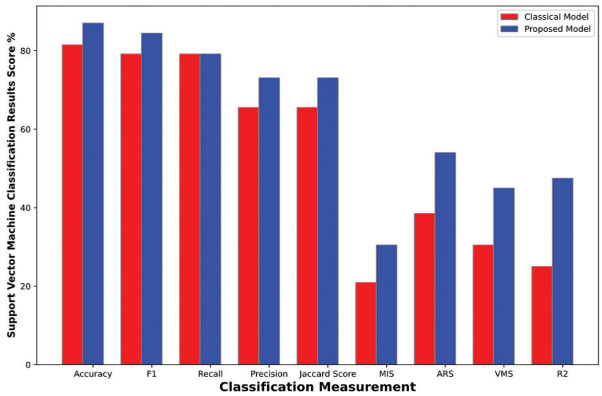

Fig. 3. Classification results assessment between classical-SVM and proposed MSMIDM based SVM algorithm

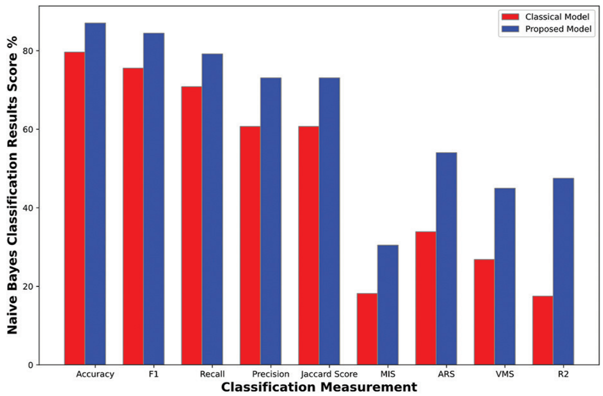

Fig. 4. Classification results assessment between classical-NB and proposed MSMIDM based NB algorithm

Fig. 5 shows that the proposed MSMIDM-based DT algorithm maximizes the confusion, statistical, and partitional-related classification results over the classical DT algorithm. The MSMIDM-based DT boosted the 2.87 % accuracy, 1.57 % F1-measure, 1.25 % recall, 2.16 % precision, 2.16 % Jaccard score, 2.16 % MIS, 3.8% ARS, 3.16 % VMS, and 6.0% R2 validation parameter as compared to the classical DT classifier.

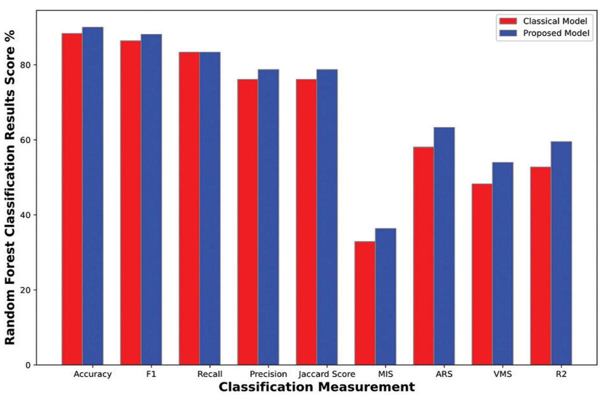

Fig. 2. Classification results assessment between classical-RF and proposed MSMIDM based RF algorithm

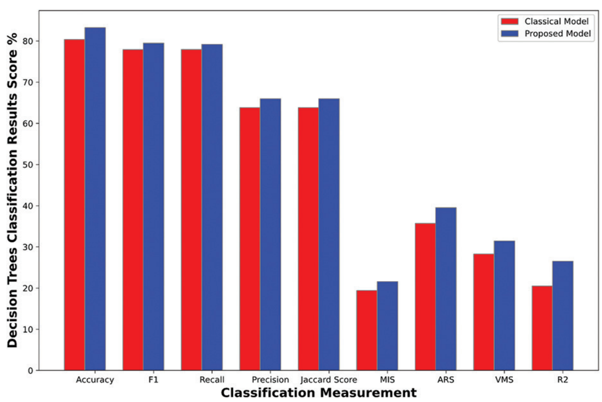

Fig. 5. Classification results assessment between classical-DT and proposed MSMIDM based DT algorithm

Fig. 6 illustrates that the MSMIDM-based KNN algorithm achieves excellent classification results compared to the classical KNN algorithm. The MSMIDM-based KNN enhanced 16.67% accuracy, 21.46% F1-measure, 65.83% recall, 24.25% precision, 24.25% Jaccard score, 17.16% MIS, 31.43% ARS, 25.45% VMS, and 20.5% R2 validation process in comparison to the classical KNN classifier.

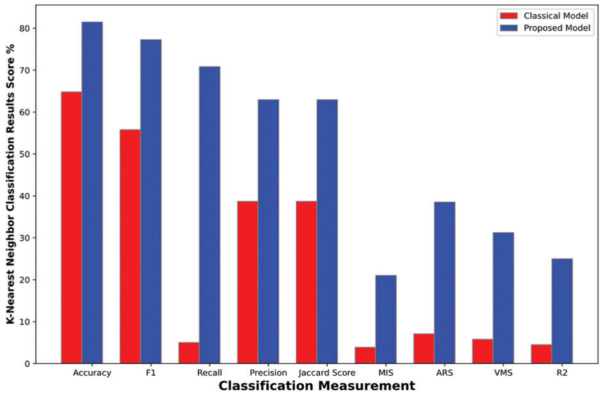

Fig. 6. Classification results assessment between classical-KNN and proposed MSMIDM based KNN algorithm

Fig. 7 demonstrates that the proposed MSMIDM-based NN algorithm obtains more robust classification results as compared to classical NN based on considered validation metrics. The MSMIDM-based NN outperformed the standard NN classifier in terms of 6.67% accuracy, 5.62% F1-measure, 0.0% recall, 7.64% precision, 7.64% Jaccard score, 6.9% MIS, 15.71% ARS, 9.58% VMS, and 27.0% R2 validation metrics.

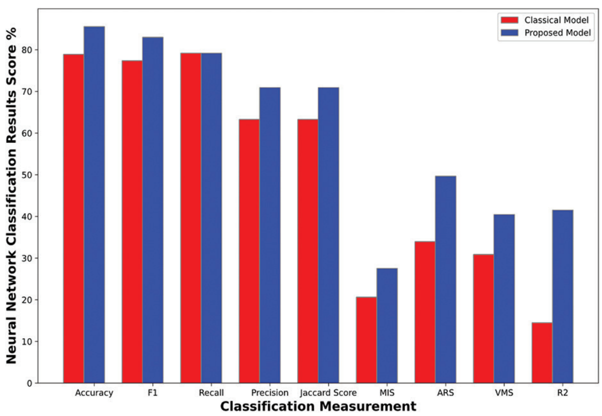

Fig. 7. Classification results assessment between classical-NN and proposed MSMIDM based NN algorithm

5. CONCLUSION

This study addresses and analyzes classification algorithms for identifying myocardial infarction heart disease. The detection of diseases is one of the tasks of the machine-learning algorithm. The classification approach extracts the hidden risk and disease-related information from the data set based on the attributes. The first section of this study describes the challenges and issues with the machine learning algorithm for detecting myocardial infarction. The second section examines various classification model literature and identifies statistical and feature selection-based approaches to improve the classification model accuracy. The selection of attributes plays a major role in disease detection. Therefore, this study uses the statistical and feature selection-based minimum skewness approach and proposes a Minimum Skewness-based Myocardial Infarction Detection Model (MSMIDM). The proposed algorithm selects the attributes based on the minimum skewness value and performs classification tasks for disease prediction through RF, NB, SVM, DT, KNN and NN classification algorithms. The experimental analysis indicates that the proposed MSMID model improves the prediction accuracy of classification algorithms. In comparison to the RF, SVM, NB, DT, KNN, and NN classification algorithms, the proposed MSMID model achieves 90%, 87.037%, 83.238%, 81.481%, and 85.556% accuracy. According to the classification results, the MSMIDM-based RF algorithm is excellent for detecting myocardial infarction and other heart diseases. Future research will use a statistical and feature selection-based approach to ensemble classification.

6. REFERENCES

[1] C. B. C. Latha, S. C. Jeeva, "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques", Informatics in Medicine Unlocked, Vol. 16, 2019.

[2] K. Uyar, A. İlhan, "Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks", Procedia Computer Science, Vol. 120, 2017, pp. 588–593.

[3] S. B. Patil, Y. S. Kumaraswamy, "Extraction of Significant Patterns from Heart Disease Warehouses for Heart Attack Prediction", International Journal of Computer Science and Network Security, Vol. 9, No. 2, 2009, p. 228.

[4] B. Fida, M. Nazir, N. Naveed, S. Akram, "Heart disease classification ensemble optimization using Genetic algorithm", Proceedings of the EEE 14th International Multitopic Conference, Karachi, Pakistan, 22-24 December 2011, pp. 19-24.

[5] A. Rairikar, V. Kulkarni, V. Sabale, H. Kale, A. Lamgunde, "Heart disease prediction using data mining techniques", Proceedings of the International Conference on Intelligent Computing and Control, Coimbatore, India, 23-24 June 2017, pp. 1-8.

[6] K. Thygesen et al. "Fourth Universal Definition of Myocardial Infarction (2018)", Circulation, Vol. 138, No. 20, 2018.

[7] V. Seenivasagam, R. Chitra, "Myocardial Infarction Detection Using Intelligent Algorithms", Neural Network World, Vol. 26, No. 1, 2016, pp. 91-110.

[8] A. Khazaee, "Heart Beat Classification Using Particle Swarm Optimization", International Journal of Intelligent Systems and Applications, Vol. 5, No. 6, 2013, pp. 25-33.

[9] D. K. Singh, V. Swaroop, "Data security and privacy in data mining: Research issues & preparation", International Journal of Computer Trends and Technology, Vol. 4, No. 2, 2013, pp. 194–200.

[10] S. M. Gorade, P. A. Deo, "A Study Some Data Mining Classification Techniques", International Journal of Modern Engineering Research, Vol. 4, No. 1, 2017, pp. 210-215.

[11] Y. Sun, J. Han, "Mining Heterogeneous Information Networks", Springer International Publishing, 2012.

[12] G. Parthiban, S. K. Srivatsa, "Applying Machine Learning Methods in Diagnosing Heart Disease for Diabetic Patients", International Journal of Applied Information Systems, Vol. 3, No. 7, 2012, pp. 25-30.

[13] L. Sun, Y. Lu, K. Yang, S. Li, "ECG analysis using multiple instance learning for myocardial infarction detection", IEEE Transactions on Biomedical Engineering, Vol. 59, No. 12, 2012, pp. 3348-3356.

[14] R. Safdari, M. Ghazisaeedi, G. Arji, M. Gharooni, M. Soraki, M. Nasiri, "A model for predicting myocardial infarction using data mining techniques", Iranian Journal of Medical Informatics, Vol. 2, No. 4, 2013.

[15] A. F. Otoom, E. E. Abdallah, Y. Kilani, A. Kefaye, M. Ashour, "Effective diagnosis and monitoring of heart disease", International Journal of Software Engineering, Technology and Applications, Vol. 9, No. 1, 2015, pp. 143-156.

[16] K. Vembandasamy, R. Sasipriya, E. Deepa, "Heart Diseases Detection Using Naive Bayes Algorithm", International Journal of Innovative Science Engineering and Technology, Vol. 2, No. 9, 2015, pp. 441-444.

[17] A. Daraei, H. Hamidi, "An efficient predictive model for myocardial infarction using cost-sensitive J48 model", Iranian Journal of Public Health, Vol. 46, No. 5, 2017, pp. 682-692.

[18] L. Verma, S. Srivastava, P. C. Negi, "A Hybrid Data Mining Model to Predict Coronary Artery Disease Cases Using Non-Invasive Clinical Data", Journal of Medical Systems, Vol. 40, No. 7, 2016.

[19] R. Chadha, S. Mayank, "Prediction of heart disease using data mining techniques", CSI Transactions on ICT, Vol. 4, No. 2-4, 2016, pp. 193-198.

[20] A. K. Dwivedi, "Performance evaluation of different machine learning techniques for prediction of heart disease", Neural Computing and Applications, Vol. 29, No. 10, 2018, pp. 685-693.

[21] S. A. Mokeddem, "A fuzzy classification model for myocardial infarction risk assessment", Applied Intelligence, Vol. 48, No. 5, 2018, pp. 1233-1250.

[22] A. Ed-Daoudy, K. Maalmi, "Performance evaluation of machine learning based big data processing framework for prediction of heart disease", Proceedings of the International Conference on Intelligent Systems and Advanced Computing Sciences, Taza, Morocco, 26-27 December 2019, pp. 1-5.

[23] S. Anitha, N. Sridevi, "Heart Disease Prediction Using Data Mining Techniques Using Data Mining Techniques", Journal of Analysis and Computation, Vol. 13, No. 2, 2019, pp. 48-55.

[24] A. K. Gárate-Escamila, A. H. El Hassani, E. Andrès, "Classification models for heart disease prediction using feature selection and PCA", Informatics in Medicine Unlocked, Vol. 19, 2020.

[25] D. Shah, S. Patel, S. K. Bharti, "Heart Disease Prediction using Machine Learning Techniques", SN Computer Science, Vol. 1, No. 6, 2020, p. 345.

[26] D. Mandair, P. Tiwari, S. Simon, K. L. Colborn, M. A. Rosenberg, "Prediction of incident myocardial infarction using machine learning applied to harmonized electronic health record data", BMC Medical Informatics and Decision Making, Vol. 20, No. 1, 2020, pp. 1-10.

[27] L. Ibrahim, M. Mesinovic, K. W. Yang, M. A. Eid, "Explainable Prediction of Acute Myocardial Infarc-

tion using Machine Learning and Shapley Values", IEEE Access, Vol. 8, 2020, pp. 210410-210417.

[28] V. N. Reddy et al. "Myocardial Infarction Prediction Using Hybrid Machine Learning Techniques", Turkish Journal of Computer and Mathematics Education, Vol. 12, No. 3, 2021, pp. 4251-4260.

[29] S. P. Patro, G. S. Nayak, N. Padhy, "Heart disease prediction by using novel optimization algorithm: A supervised learning prospective", Informatics in Medicine Unlocked, Vol. 26, 2021.

[30] A. Kondababu, V. Siddhartha, B. B. Kumar, B. Penumutchi, "A comparative study on machine learning based heart disease prediction", Materials Today: Proceedings, 2021, pp. 1-5.

[31] R. Liu et al. "An artificial intelligence-based risk prediction model of myocardial infarction", BMC Bioinformatics, Vol. 23, No. 1, 2022, pp. 1-17.

[32] U. Nagavelli, D. Samanta, P. Chakraborty, "Machine Learning Technology-Based Heart Disease Detection Models", Journal of Healthcare Engineering, Vol. 2022, 2022.

[33] P. A. Sadiyamole, S. M. Priya, "A Robust Cardiovascular Disease Predictor Based on Genetic Feature Selection and Ensemble Learning Classification", International Journal of Electrical and Computer Engineering Systems, Vol. 14, No. 7, 2023, pp. 799-809.

[34] A. A. Ahmad, H. Polat, "Prediction of Heart Disease Based on Machine Learning Using Jellyfish Optimization Algorithm", Diagnostics, Vol. 13, No. 14, 2023.

[35] C. M. Bhatt, P. Patel, T. Ghetia, P. L. Mazzeo, "Effective Heart Disease Prediction Using Machine Learning Techniques", Algorithms, Vol. 16, No. 2, 2023, p. 88.

[36] S. Subramani et al. "Cardiovascular diseases prediction by machine learning incorporation with deep learning", Frontiers in Medicine, Vol. 10, No. April, 2023, pp. 1-9.

[37] H. A. Al-Alshaikh et al. "Comprehensive evaluation and performance analysis of machine learning in heart disease prediction", Scientific Reports, Vol. 14, No. 1, 2024, pp. 1-15.

[38] X. Zhao, J. Liang, C. Dang, "A stratified sampling based clustering algorithm for large-scale data", Knowledge-Based Systems, Vol. 163, 2019, pp. 416-428.

[39] S. I. Ayon, M. M. Islam, M. R. Hossain, "Coronary Artery Heart Disease Prediction: A Comparative Study of Computational Intelligence Techniques", IETE Journal of Research, Vol. 68, No. 4, 2022, pp. 2488-2507.

[40] S. Mohan, C. Thirumalai, G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques", IEEE Access, Vol. 7, 2019, pp. 81542-81554.

[41] R. Katarya, S. K. Meena, "Machine Learning Techniques for Heart Disease Prediction: A Comparative Study and Analysis", Health and Technology, Vol. 11, No. 1, 2021, pp. 87-97.

[42] J. P. Verma, "Normal Distribution and Its Application", Statistics and Research Methods in Psychology with Excel, Springer, 2019, pp. 201-235.

[43] S. K. Yadav, S. Singh, R. Gupta, "Measures of Dispersion", Biomedical Statistics, Springer, 2019, pp. 59-70.

[44] K. K. Pandey, D. Shukla, "Min–max kurtosis mean distance based k-means initial centroid initialization method for big genomic data clustering", Evolutionary Intelligence, Vol. 16, 2022, pp. 1055-1076.

[45] K. K. Pandey, D. Shukla, "Maxmin distance sort heuristic-based initial centroid method of partitional clustering for big data mining", Pattern Analysis and Applications, Vol. 25, No. 1, 2022, pp. 139-156.