[1] Nudrat Fatima *

[2] Dr. Sifatullah Siddiqi

Enhanced Myocardial Infarction Prediction Using Machine Learning Algorithms and Gender-Specific Insights

Journal of Electrical Systems

Abstract: - A serious medical condition called myocardial infarction (MI), sometimes referred to as a "heart attack," is caused by disruptions in the blood supply to the myocardium. This research examines the efficacy of machine learning (ML) algorithms in forecasting myocardial infarction (MI) using a dataset of 350 records. The study identifies key risk factors for predicting myocardial infarction (MI), such as elevated cholesterol levels, diabetes, advanced age, overall health status, mental well-being, obesity, physical activity, smoking habits, hypertension, and depression. Significantly, gender does not manifest as a predictor of myocardial infarction (MI) when employing various classification methods. The research achieves high accuracy rates of 89.32%, 87.53%, 81.29%, and 76.59% using different machine learning algorithms, including Deep Belief Network (DBN), C4.5, Random Forest (RF), and Bayesian Network (BN), respectively. Algorithm-specific rule sets identify correlations, with the C4.5 algorithm revealing interesting connections between smoking habits and protection against myocardial infarction (MI). Performance metrics like accuracy, precision, sensitivity, and specificity attest to the effectiveness of the proposed technique. The results demonstrate the superior performance of the DBN algorithm, surpassing other algorithms in terms of accuracy (89.32%), precision (84.04%), sensitivity (86.63%), and specificity (82.45%). This paper provides crucial insights into predictive modeling for myocardial infarction (MI), highlighting the importance of various risk factors and advanced machine learning (ML) algorithms. The results offer clinicians and researchers a strong foundation for comprehending and potentially averting myocardial infarction, relying on personalized patient profiles. This paper has the potential to significantly contribute to the field by applying ensemble classifiers and machine learning models to forecast gender-specific myocardial infarctions. As a result, diagnostic precision and patient outcomes could be revolutionized.

Keywords: Machine Learning, Ensemble Classifier, Myocardial infarction, Random Forest, Multi-layer perceptron, Disease Prediction, Prediction Algorithm, Medical Diagnosis, Deep Believe Network.

I. INTRODUCTION

In a global landscape where Every year, information technology becomes more prevalent in all sectors of activity. The advancement of computer information technologies, which have been widely used in recent decades, has also embraced medicine and has been increasing speed. A wide range of sectors has seen radical change because of the global Coronavirus Disease (COVID-19) pandemic's unprecedented acceleration of digitalization [1]. Information systems are being utilized increasingly in medicine nowadays, from diagnosing patients to predicting the resources required for the ongoing functioning of medical facilities [2].

One of the most prevalent CVD in the trauma sector is acute MI [3]. MI happens when cardiac tissue dies because of myocardial distress. The most frequent cause of thrombotic blockage of a coronary artery in the clinical setting is the breakdown of an atherosclerotic plaque. Ischaemia causes disturbances in the MI, which causes a quick decline in cardiac output. If ischemia persists for a long time, myocardial necrosis could also actually occur. If revascularization is postponed or fails, the severe damage can cause PMO (Persistent Microvascular Obstruction), also referred to as the no-reflow effect [4]. Therefore, it is essential to begin emergency revascularization treatment as soon as the condition is identified to restore perfusion.

Acute Myocardial Infarction (AMI)—a heart attack—is one of the worst CVD diseases. When the heart's muscles lose blood flow, it damages them. It is also crucial to correctly diagnose heart disease (HD). HD must be diagnosed using the symptoms, physical exam, and knowledge of the many indicators of this disease. Many factors contribute to HD development and progression, such as high cholesterol levels, family history of heart disease, hypertension, inactivity, obesity, and tobacco use. The principal cause of myocardial infarction is the obstruction of blood flow to the coronary arteries. Reduced blood flow leads to a reduction in Red Blood Cells (RBC), which prevents the body from obtaining sufficient oxygen and results in unconsciousness. An early diagnosis using symptoms and indicators can help patients avoid heart attacks If the prediction is reliable enough [5]. Different heart attack symptoms are shown in Figure 1. The primary factor for heart attacks is coronary artery disease (CAD).

[1] Scholar, Computer science and Engineering, Integral University, Lucknow, Uttar Pradesh-226026, India.

[2] Computer science and Engineering, Integral University, Lucknow, Uttar Pradesh -226026, India. sifatullah.siddiqi@gmail.com

* Corresponding Author Email: nudratfatima2@gmail.com

Older patients (more than 55 years for women and more than 45 years for males) or who have a prior familial history of early-stage coronary artery disease are also at an increased risk of developing plaque.

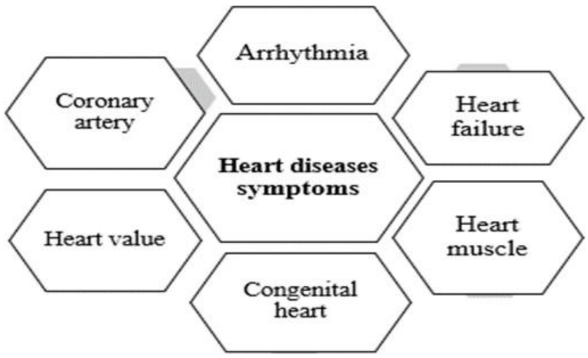


Figure 1. Symptoms of heart attack [6].

Preventing CVD could be as simple as educating people about their risk factors for CAD and guiding them to take preventive action based on that information [7]. Early and precise identification could result in improved decision-making, as is the case with many complicated medical issues. A crucial factor to consider while developing a CAD diagnostic model is the patient's gender. The development of separate CAD diagnostic models for women and men expedites the detection, treatment, and decision-making process. Gender-based analysis affects every stage of the diagnostic model creation process [8] [9].

A technique for extracting and manipulating implicit, unknown, or known but extracting meaningful insights from datasets called ML [10]. The area of ML is very broad and diversified, its scope and utility are ever-expanding. ML incorporates several classifiers from Supervised Learning (SL), Unsupervised Learning (UL), and Ensemble Learning (EL) that are used to forecast and determine the correctness of the provided dataset [11]. Ensemble learning is a robust method for generating a classifier that is close to optimal for any given problem. By strengthening the ensemble in subspaces prone to errors, this approach has the potential to enhance classification performance. In general, it is accurate to state that improved classification is the outcome of combining a variety of classifiers. Developing a classifier ensemble presents a primary obstacle in the form of establishing a general strategy to guarantee diversity, an essential element of any ensemble [12].

The authors want to evaluate the effectiveness of ML techniques over more conventional statistical methods in this situation. High blood pressure, high cholesterol, and tobacco use are the key risk factors for cardiovascular disease. Smoking status was shown to be the strongest predictor of MI across all four algorithms used in this study. The study follows the following structure: In Section 2, the authors provide a literature assessment of gender-based prediction of MI using ML. In Section 3, the authors survey prior study on the prediction of MI, and Section 4, states the issue and provide a solution. Section 5 defines the research objective. The study methodology and methods are introduced in Section 6, and the recommended method is proven. The latter section is the basis for the anticipated findings presented in Section 7 Finally, the report makes a conclusion and suggests areas for future research.

A. Myocardial Infarction

Coronary heart disease, also referred to as a heart attack, affects the blood vessels that supply the heart muscle (myocardium). A MI occurs when a segment of the heart muscle receives no blood supply or too little to maintain cardiac muscle activity [13] [14]. Figures 2 and 3 show the heart with MI and the categorization of MI, respectively. The categories for MI are:

- **NSTEMI:** NSTEMI referred as Non-ST-Elevation Myocardial Infarction. It describes a kind of acute coronary syndrome that includes three conditions that reduce cardiac blood flow.
- **STEMI:** When myocardial damage or necrosis is caused by transmural ischemia, this condition is called as acute STEMI (ST-Elevation Myocardial Infarction) [15].

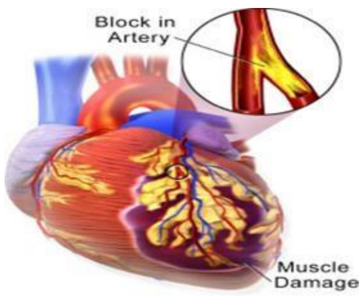


Figure 2. Heart with MI [16].

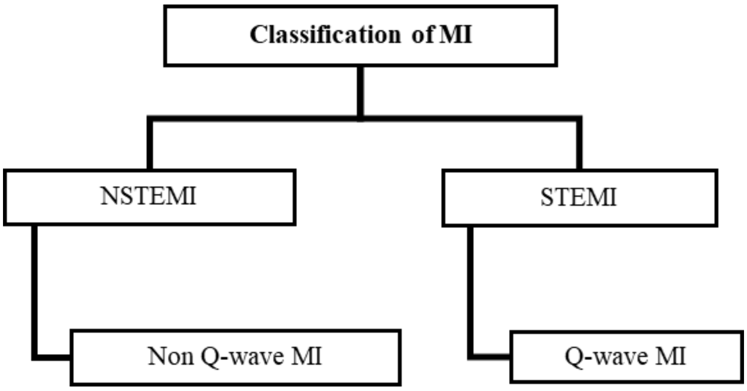


Figure 3. Classification of MI

The likelihood of suffering a heart attack depends on a variety of factors. Certain heart attack risk factors are unfortunately out of the hands [17]. Table 1 indicates certain risk factors of heart attacks. It explains why each factor is considered as a risk and gives the additional detail of the factor.

Table 1. Coronary artery disease risk factor

Coronary artery disease risk factor	Why it is a risk factor	Details
Age and sex	Age increases heart attack risk. Sex affects heart attack onset.	People and men AMAB: Heart attacks increase dramatically at 45. People and women AFAB: Heart attacks increase after menopause or about 50. AMAB- means born male. AFAB—Assigned Female at Birth.
Family history of heart disease	The risk is increased if a parent or sibling has had a heart attack or has had one in the past, particularly if they were younger because the genetic makeup is like theirs.	The risk increases if: The chance of developing HD is increased if either the father or a sibling was diagnosed with the condition before the age of 55. Mother or sibling had HD at 65 or younger.
Lifestyle	Making heart-unhealthy lifestyle decisions can make it more likely to have a heart attack.	These increase heart attack risk: No exercise. A high-fat, sugar, and salt diet. Tobacco use. Alcohol abuse. Drug misuse in youth.

II. Literature Review

In this section, the author defines the Literature review on the gender-based prediction of myocardial infarction (MI) using machine learning (ML).

Han et al., (2023) [18] developed a model called MLB-ResNet-SENet-BL for bidirectional long-term short- term memory (LSTM). The leading morphological information representation network, which is built on MLB-ResNet, initially employed geographical characteristics. Afterwards, the central network for feature mapping significance analysis adjusted the strength of the SENet-based feature mappings between these geographical features. The final model achieves better results than the most recent cutting-edge investigations in both intra- and inter-patient settings. Grad-CAM, an interpretability analysis based on class activation mapping, allowed for the visualization of the ST-T segments and QRS waves in 12-lead electrocardiograms. The highlighted areas of the heat maps aligned perfectly with the diagnostic foundation and approach utilized by doctors, as demonstrated by this. The implementation of such models has the ability to improve treatment outcomes while also protecting patients' lives.

Xiong et al., (2022) [19] suggested a deep learning method for locating and detecting myocardial infarction via electrocardiogram. In the past five years, 59 major DL studies have been published that apply ECG localization and detection of myocardial infarction (MI). These studies have utilized LSTM (long short-term memory), ResNet (residual neural network), CRNN (convolutional recurrent neural network), AE (autoencoder), and GRU (gated recurrent unit). During this time period, CNN achieved the greatest level of popularity for both localization and detection of MI, and the ResNet model and CNN achieved the highest performance. The atmost accuracies reported for each of the six distinct methods exceed 97%. When contrasting the utilization of various datasets and ECG leads, it is observed that the precision of the network fed 12-lead electrocardiogram data from the PTB database is superior to that derived from smaller lead counts in other datasets.

Bat-Erdene et al., (2022) [20] suggested a Deep Learning (DL)-based model for predicting readmission to the hospital for Heart Failure (HF) 6, 12, and 24 months after AMI patients have been released from the hospital AMI. As compared to more conventional ML techniques including gradient boosting machine, Random Forest (RF), AdaBoost, Logistic Regression (LR), Support Vector Machine (SVM), the suggested DL-based rehospitalization prediction model performed very well. The suggested model achieved results of 99.37% accuracy, 99.49% specificity, 99.90% Area Under the Curve (AUC), 98.61% recall, 96.86% precision, and 97.73% F1 score.

Xue et al., (2021) [21] introduced a new ML technique to distinguish between three phenogroups of STEMI patients who had distinct lipoprotein characteristics. The whole patient population was divided into phenotypes 1 through 3. Phenogroup 1 patient was classified as having a statin-modified cardiovascular risk due to their high Lp
(a) and low HDL-C and apoA1 values. Phenotype 2 patients had the highest HDL-C levels as well as apoA1 levels and the lowest TG, TC, LDL, and apoB values. TG, TC, LDL-C, and apo B were highest in phenogroup 3, while Lp(a) was lowest. Furthermore, the prognosis was worse for those in phenogroup 1. Patients in phenogroup 1 had a considerably greater risk for all bad outcomes, as shown by a multivariate Cox analysis.

Bai et al., (2021) [22] evaluate the performance of several ML models in forecasting the 12-month risk of death for STEMI patients exhibiting hyperuricemia. Five ML models were evaluated against the conventional global (GRACE) risk score for acute coronary event reports LR, XG-Boost, RF, Cat-Boost, and K-nearest neighbor. The GRACE risk model was one of six with an AUC between 0.75 and 0.88. Cat-Boost outpaced the other models in terms of AUC (0.87), precision (0.84), prediction accuracy (0.89), and F1 value (0.44). Cat-Boost got the greatest accuracy (0.96), precision (0.95), AUC (0.99), and F1 value (0.97) after the hybrid sampling approach modification.

Li et al., (2021) [23] examined the connection between levels of plasma D-dimer and subsequent hospitalization for HF in patients with STEMI who had had primary Percutaneous Coronary Intervention (pPCI). Baseline D-dimer levels were compared among males and females, and participants were separated into two groups based on their plasma D-dimer levels. The connection between D-dimer and HF hospitalization was analyzed using LR and Receiver Operating Characteristic (ROC) curves. D-dimer was shown to be an individual predictor of hospital-acquired HF in both the general population and female patients using a multivariate LR model.

Liang et al., (2020) [24] developed two groups of patients with HF after MI with median ages of 71 and corresponding prediction models for the two age groups, respectively. To build prediction models, three distinct ML techniques such as LR, RF, and XG-Boost were utilized. A 5-fold cross-validation was then accomplished to assess the predictive accuracy and stability of the prediction models. Analytic results from three different ML techniques were consistent in suggesting that age stratification preceded data training for improved system performance when building prediction models of MI leading to accelerated transition towards HF within a specific interval.

Kamalapurkar et al., (2020) [25] suggested study aims to improve the accuracy with which coronary disease can be predicted in a patient. The study presents a web-based system that uses ML algorithms to predict cardiovascular illness with high accuracy compared to previous efforts. It employs an ensemble classification
approach for the prediction of heart illness since ensemble methods offer superior accuracy in comparison to individual classifiers like SVM or RF.

Kayyum et al., (2020) [26] suggestions for ML methods for data gathering and categorization. It has 345 cases and 26 characteristics gathered. There are three distinct classifications available in the class attribute: Distinctive, Non-Distinctive, and Both. The training of the dataset has been done utilizing K-Fold Cross Validation Technique and particularly three ML algorithms have been utilized which are Bagging, LR, and RF. Moreover, the study has the potential to demonstrate that the ML algorithms have an accuracy of 93.913%, 93.6323%, and 91.0145%, respectively.

Than et al., (2019) [27] suggested that changes in cardiac troponin concentrations among age groups, sex, and testing windows in people alleged of having Myocardial Infarction are not used in present for diagnostic practices. An ML approach was created for 3,013 persons by including age, gender, or an amalgamation of these parameters with very sensitive cardiac troponin I value, and it was evaluated on 7,998 patients with suspected myocardial-ischemic-injury-index 〖MI〗^3. The criteria used in 〖MI〗^3's training sets to categorize impacted roles as high-risk (75% PPV) and low-risk (99% sensitivity) were evaluated in separate research for their applicability in identifying potentially hazardous detours.

Liu et al., (2018) [28] proposed of a novel MFB-CNN, an ECG-based Multiple-Feature-Branch Convolutional Neural Network, is suggested to detect and localize myocardial infarction automatically. Every individual autonomous characteristic branch of the MFB-CNN is associated with a distinct lead. A feature branch can acquire knowledge of specific attributes of a lead by capitalizing on the variety present among the twelve leads. The integrity can be exploited by a globally connected fully-connected softmax layer that describe of every feature branch. In order to assess the proposed algorithm, ECG (electrocardiogram) data from the PTB diagnostic database is utilized.

It is capable of performing admirably in the diagnosis of MI. The maximum average accuracies to facilitate the detection and localization of MI based on class are 99.81% and 99.95%, respectively; for patient- specific experiments, those figures are 94.82% and 98.79%, respectively.

Diker et al., (2017) [29] suggested a diagnostic application model for the classification of electrocardiogram (ECG) signals. This model is constructed using a combination of two machine learning technique, ANN (artificial neural networks) and a-nearest neighbors (a-NN) and incorporates a Recursive Feature Eliminator (RFE). For the experiments, an open-access ECG database was utilized. Prior to that, the signals underwent pre-processing. From ECG signals, a number of diagnostic characteristics extracted from the statistical and morphological domains. In the final phase of the analysis, myocardial infarction (MI) samples were distinguished from normal using the aforementioned machine learning techniques and the RFE algorithm with 10-fold cross-validation. Achieving an sensitivity of 86.58%, specificity of 64.71%, and accuracy of 80.60% yielded encouraging outcomes.

Waqar et al., (2021) [30] proposed research that provides a dependable and cost-effective method for predicting myocardial infarction. It predicts the occurrence of a myocardial infarction using multiple machine learning algorithms and a UCI dataset, with no feature engineering involved. In addition, the positive and negative classifications in the provided dataset are not distributed equally, which may hinder performance. The proposed work handles imbalanced data using a synthetic minority oversampling technique (SMOTE). The necessity for feature engineering in the classification of the provided dataset was eliminated by the proposed system. This resulted in an effective resolution, given that feature engineering is frequently an expensive endeavor. The findings indicate that, when appropriately calibrated, an SMOTE-based artificial neural network exhibited superior performance compared to numerous pre-existing systems and alternative machine learning algorithms. The efficacy of the proposed system in heart attack prediction is guaranteed by its exceptional dependability.

III. Problem Formulation

Myocardial infarction, angina pectoris, and hyperlipidemia are all examples of cardiovascular disease. Electrocardiography, ultrasonography, angiography, blood tests, and other methods are used to identify CVD. Because such techniques necessitate several testing’s, they are expensive and time-consuming. ML-based CVD prediction techniques have recently been created to improve existing diagnostic techniques. ML is a sort of AI used to handle a wide range of data science problems. ML is frequently used to predict a result based on previously collected data. The classifier is a widely used ML technique for predicting outcomes. Some classification algorithms predict with a high level of accuracy, while others have a lower level of accuracy. In the literature, the calculation of the CVD member classification and prediction techniques are rated and applied. The major concern is the specific techniques well for defined circumstances and specific to the dataset. There is a need to have a common technique that can work with almost all types of considerations and stats for the validation of the selection of the prediction model. In the current work, 5 classification techniques such as RF, NBC, C4.5, BN, and DBN are applied on separate dataset sub-model, and the ensembled prediction and individual prediction for each technique are counted for selection as optimized prediction techniques. This validation uses the U test and chi-square for gender, age, SBP, DBP, cholesterol levels, diabetes, and smoking variables to create the optimum prediction model. Most researchers [22], [26] have worked out how machine learning models could be used alone or with a narrow focus on certain demographic characteristics, but there is still some gap in the possible synergies that could happen when traditional risk factors are combined with advanced predictive algorithms. Having reviewed the research papers on heart attacks, a combination of conventional risk factors—age, sex, family history, and lifestyle—influence the identification of a heart attack. This study's goal is to investigate the combination of machine learning models and traditional risk factors to construct a predictive framework for heart attacks that is more resilient and precise. There is still an area of research that has yet to be explored, which is the integration of demographic and lifestyle risk factors with machine learning models to improve the accuracy of predictions following heart attack risk factors. Further research is warranted to determine how the incorporation of lifestyle choices, age, sex, and familial predisposition to heart disease can enhance the accuracy of machine learning-based heart attack predictions.

IV. Research Methodology

In this research study, two methods of ML are used for the prediction of Myocardial Infarction: (i) Deep Belief Network (DBN) and (ii) Ensemble Classifier.

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

A. Techniques Used

- **Deep Belief Network (DBN)**

There are primarily two distinctions between Deep Neural Networks (DNN) and DBN:

➢ Network topology: A DNN has several hidden layers and is a feedforward network. The logistic/sigmoid activation function is often used by every hidden neuron. In contrast, the DBN hidden layers of stacked Restricted Boltzmann Machines (RBM) are connected randomly.

➢ Network training: Backpropagation training of a DNN relies on labeled data throughout to fine-tune its weights. DBN, on the other hand, conducts its initial training in an unsupervised manner through contrasting divergence and then fine-tunes its weights using backpropagation.

DNNs need a large quantity of evenly distributed labels; however, most real-world datasets lack such labels. A DBN is a probabilistic generative model that is often built by layering RBM. The Contrastive Divergence (CD) approach is used to fine-tune the parameters of stacked RBM [31]. The CD is an unsupervised kind of learning; hence labeled data are not necessary at this point. Afterward, the pre-trained network would be fine-tuned using an SL model, such as SoftMax/LR or a linear classifier, and the gradient descent learning process [32]. However, following CD, the DBN parameters are mostly set, and the second stage only adjusts the model's features. Therefore, DBN requires a less amount of data to be tagged [33]. Figure 4 depicts a typical organizational layout for a network [34].

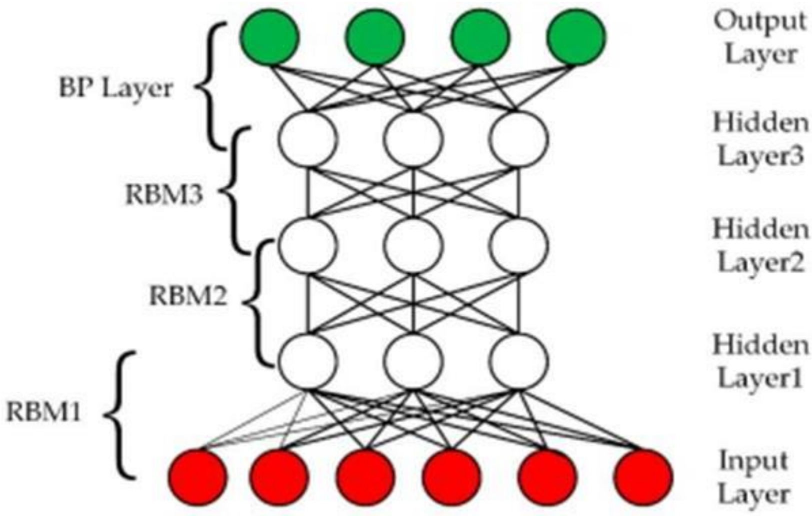


Figure 4. DBN basic network structure [34]

B. Ensemble Classifier

Ensembles improve classifier accuracy. It connects weak learners with strong learners to improve their performance. Ensemble methods improve heart disease prediction systems.

The goal of integrating several classifiers is to achieve greater performance than a single classifier [35]. The process for the ensemble is illustrated in Figure 5.

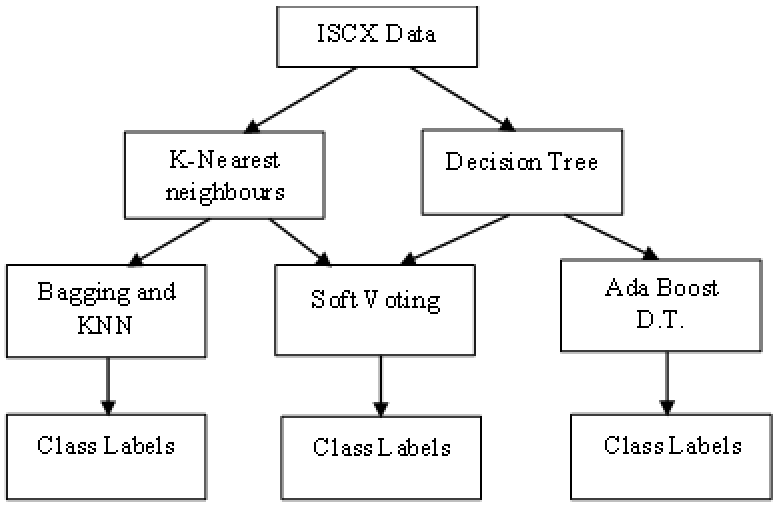


Figure 5. Ensemble classifier methods [36] The different classifier that is used in the process are:

- **Bayesian Network (BN)**

The primary theory of the Bayesian Network classifier is that the characteristics are individual to one another. The opposite extreme theory is that all the qualities are dependent on one another. A directed acyclic graph with random variables as nodes and conditional dependencies as edges are the result of the BN model. Models like this are called BN. Complete models for the variables and their connections are regarded as BNs [37].

In this application of BN, the author’s state of mind is characterized by three distinct (but interconnected) objectives, all of which were considered. They proposed a classification strategy based on multi-dimensional BN classifiers. It grouped all the target variables into one classification task, capitalizing on potential relationships between them. To make use of the vast amount of unlabeled data available from a semi-supervised setting, it was chosen to extend the multi-dimensional classification framework to that area. In a semi-supervised setting, their categorization best reflects the genuine underlying domain structure [38].

- **Naïve Bayes Classifier (NBC)**

The Bayesian classifier, often identified as the NBC, is cantered on Bayes' theorem. One method of text categorization is allocated to a certain document d the class c*= arg $maxc\ p(c|d)$.

It develops the NB classifier by first reflecting that by Bayes' rule,

$$P(c|d) = \frac{P(c)P(d|c)}{P(d)} \quad (1)$$

where P(d) is not relevant in selecting c^*. To approximate the term P (d | c), NBC it by supposing the $f'_i s$ are provisionally individually given d's class:

$$P_{NB}(c|d) := \frac{P(c)(\prod_{i=1}^{m} P(f_i|c)^{n_i(d)})}{P(d)} \quad (2)$$

The training process comprise of relative-frequency estimation of $P(c)$ and $P(fi|c)$, employing add-one smoothing. However simple Naive Bayes-based text categorization is, it works quite well, even though its provisional objectivity assumption does not apply in real-world contexts [39].

- **Random Forest (RF)**

This is one of the ensemble approaches that is exclusively utilized to increase the success and accuracy of ML algorithms in artificial intelligence. An RF technique could also aid in identifying the relevant independent variables and allowing the system to choose functionality. The relevance of picking many options for each shrub in the empirical investigation has already been established by several discoveries, and this method is shown to be ideal in terms of prediction accuracy [40]. The Decision Tree (DT) with classification model and regression model for dependent variables is shown in Figure 6.

Scikit-learn uses Gini Significance to estimate the importance of each node in a DT, given that there are just two offspring/child nodes.

$$nij = wj\ Cj - wleft(j)Cleft(j) - wright(j)\ Cright(j) \quad (3)$$

Where, nij = the importance of node j

wj = weighted number of samples reaching node j

Cj = the impurity value of node j

$left(j)$ = child node from left split on node j

$right(j)$ = child node from right split on node j

Where, nij represents the significance of node j, wj is the weighted number of samples reaching node j, Cj represents the impurity value of node j, $left(j)$ shows child node from left split on node j and $right(j)$ is child node from right split on node j.

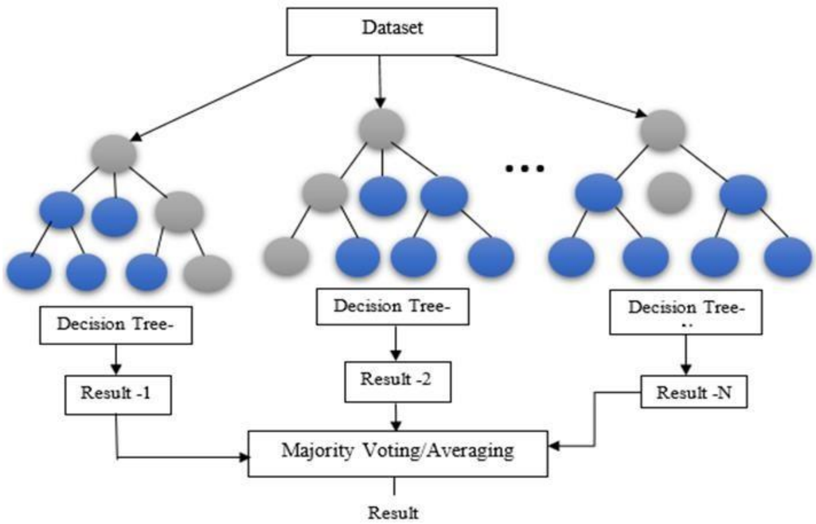


Figure 6. The structure of RF [41].

- **C4.5**

The ID3 algorithm, which is a simple DT algorithm, provides the basis for the C4.5 algorithm. Quinlan produced the algorithm. It splits the trees using the information gain ratio as a metric. It takes data as input and outputs a DT as the result. Univariate trees are created using this approach. DT is used to represent classification rules. When the split of a tree falls below a particular threshold value, it is halted. It prunes based on errors and is an excellent approach for dealing with numeric properties.

- **Multilayer Perceptron (MP)**

One of the primary limitations of the perceptron was its ability to solve problems that could be separated linearly. Relying on units with a threshold activation function, the basic Perceptron was limited to resolving linearly separable problems.
However, numerous complex problems in artificial intelligence (AI) lack linear separability. As a result, a critical vulnerability in the Perceptron was identified. To return to the issue of modelling logic gates, it was found that the exclusive-or problem (XOR) is also not linearly separable [42].
The multilayered perceptron is the form of neural network that is most widely recognized and utilized. Training a multilayer perceptron model frequently incurs significant latency, with complex problems necessitating tens of thousands or even thousands of epochs [43].

4.1. Some Common Mistakes

The proposed method in this research validates the result for gender-based prediction of MI by DBN and Ensemble Classifier technique. The proposed methodology is depicted in Figure 7.

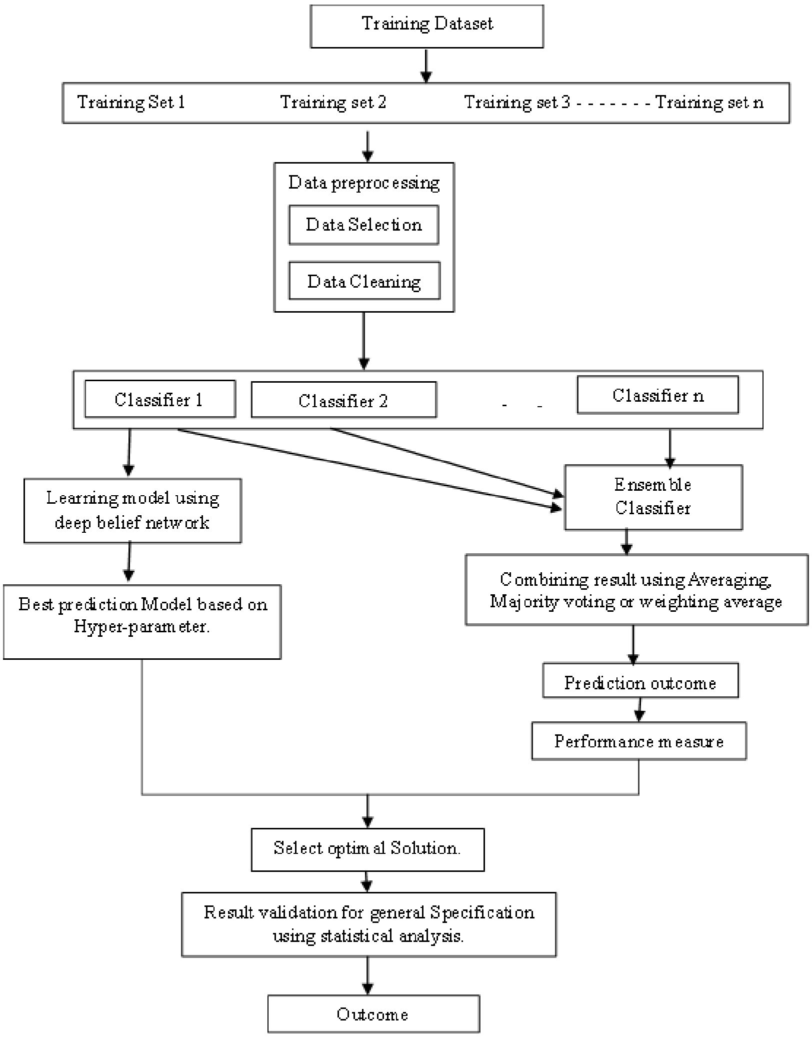


Figure 7. Proposed methodology

Below are the steps for the data process for Gender-Based Prediction: -
Step 1: Data set collection and training

Dataset collection is done and after that training of the dataset is done according to the ML that is applied to predict the outcome. This training data is divided into equal subtasks based on the requirements.

Step 2: Splitting of data

After splitting the training dataset into equal parts data pre-processing is taken out. This data pre-processing contains the selection of data and data cleaning. In the selection process, a specific type of dataset is considered that is required for the analysis. And the cleaning of the dataset is carried out that contains noise removal (unnecessary data deletion) from the dataset and the removal of repeated data.

Step 3: Data Classification

After all the process of data selection and cleaning is done then the data classification is carried out (A classifier is a controlled learning process for predicting outcomes from previously collected data). With the classifier, two techniques are proposed.

- Deep Belief Network
- Ensemble Classifier.

Step 4: Data after classification is then transferred to DBN and Ensemble Classifier.

- Data from the learning model using the DBN process for the hyper-parameter (model based on best prediction)
- Similarly, from the classifiers, the data is processed to the Ensemble classifier technique, and with the use of weight averaging and major voting of data, the result is combined and predicted outcomes performance is measured.

Step 5: Result Optimization

With both the outcomes taken from the solution of the selected techniques, the data is optimized. Finally, the proposed model gives the prediction of the MI based on gender.

V. Result And Implementation

In this part, the author would demonstrate how the recommended strategies can be implemented.

A. Dataset

There was a total of 350 records in the dataset (200 MI patients and 150 controls) and 40 individual variables. Those with whom the occurrence of MI had been denied based on evidence in their medical records were referred to as "people without MI" in the research. Patients are patients whose medical records have been used to confirm the presence of MI.

- **Prediction of MI**

High cholesterol, diabetes, age, general health, mental health, obesity, exercise, cigarette use, High blood pressure, and depression are all recognized to be behavioural and risk factors (comorbidities) for a MI [44-45]. Risk factors for MI include poor self-reported health status [46], smoking [47], high cholesterol [48], high blood pressure [49], days physical health bad not excellent, stroke, chronic obstructive pulmonary disorder-COPD, renal illness, and arthritis. Moderate alcohol use has also been proven to minimize the risk of MI, but heavy drinking significantly raises the risk. Mental health and depressive disorders are psychological aspects connected to having a MI. The significance of these variables for the categorization of MI is shown in Figure 8. Figure 8 shows the variables that are important in the classification of MI. Gain estimates of how well a particular attribute impacts the prediction of an ML model.

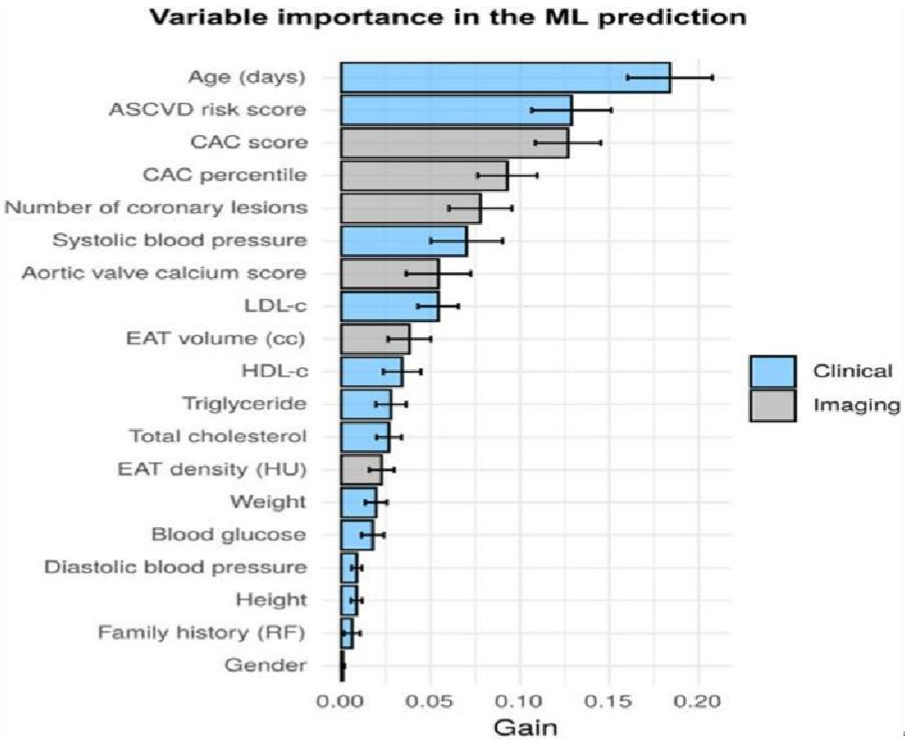


Figure 8. Importance of Variables in the Classification of MI. Gain quantifies how favorably a given attribute influences an ML model's prediction

There are a total of 129 patient records (47% of the total) and 145 non-MI person records (53% of the total) included in the analysis. Most of the information was collected from guys (58%). Patients and healthy individuals had an average (SD) age of 60 and 57 years, respectively. In Table 3 authors provide the most fundamental data about the study's patients. Examine Figure 9 for a chart depicting the sex distribution of MI cases and controls.

Table 3. Participants' demographic information

	People without MI	**People with MI**
Age		
Mean	60	57
Std. Deviation	11.12	10.66
Min	30	33
Max	85	91
Gender		
Male	101 (37%)	58 (21%)
Female	44 (16%)	71 (26%)
Frequency	145 (53%)	129 (47%)

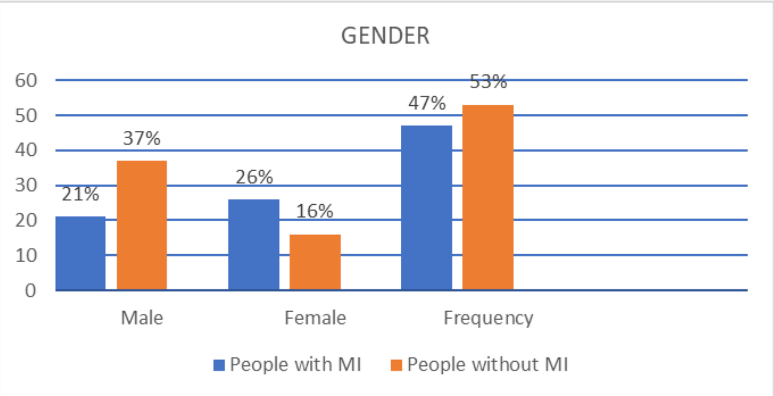


Figure 9. Graph of gender-based MI people

DBN, ensemble classifier (C4.5, RF, BN) methods were able to predict MI with 89.32%, 87.53%, 81.29%, and 76.59% accuracy, respectively [Table 4]; these results were the best of the modelling strategies utilized. The C5 algorithm's resulting set of rules can be found in [Table 5]. Figure 10 indicates the graph of proposed methods based on accuracy, precision, sensitivity, and specificity.

The following performance metrics are used to evaluate the effectiveness of the suggested technique.

- $Total\_Images$ The sum amount of examined pictures.
- TP (True positive): Detected the altered images without error.
- TN (True negative): Validated as authentic on visual inspection.
- FN (False negative): falsely recognized manipulated images or manipulated images that were mistakenly thought to be genuine.
- FP (False positive): images that have been mistakenly recognized as being genuine or as being manipulated. To evaluate the suggested technique and to make comparisons to others, the authors compute its accuracy,

precision, recall, and fallout. The formulas for these are as follows

$$Accuracy = \frac{T_P+T_N}{T_{Total\_Images}} \times 100 \tag{4}$$

$$Precission = \frac{T_P}{T_P+F_P} \tag{5}$$

$$Recall = \left(\frac{T_P}{T_P+F_N}\right) \tag{6}$$

Table 4. The algorithms were compared for their accuracy, precision, sensitivity, and specificity.

Techniques	Accuracy	Precision	Sensitivity	Specificity
DBN	89.32%	84.04%	86.63%	82.45%
C4.5	87.53%	82.75%	83.69%	80.37%
RF	81.29%	79.45%	80.58%	78.35%
BN	76.59%	72.83%	77.29%	74.33%

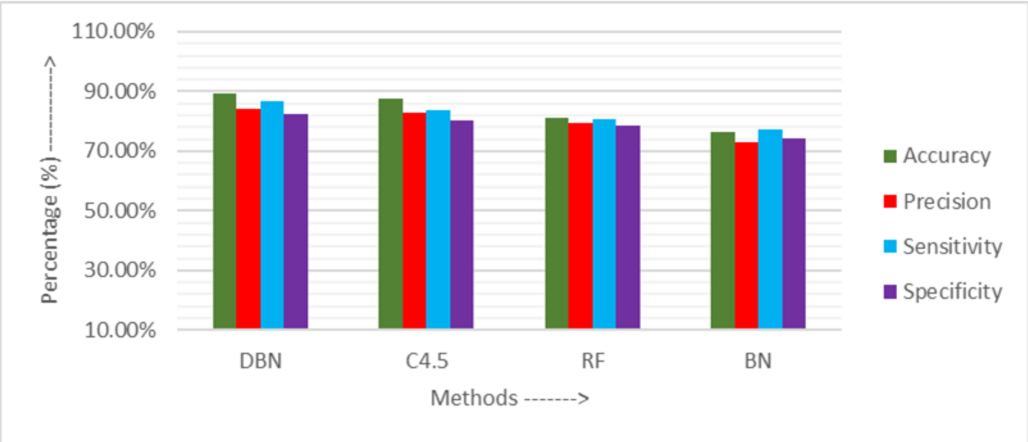


Figure 10. Graph of the proposed method

Table 5. C4.5-based rule sets and their respective accuracy percentages

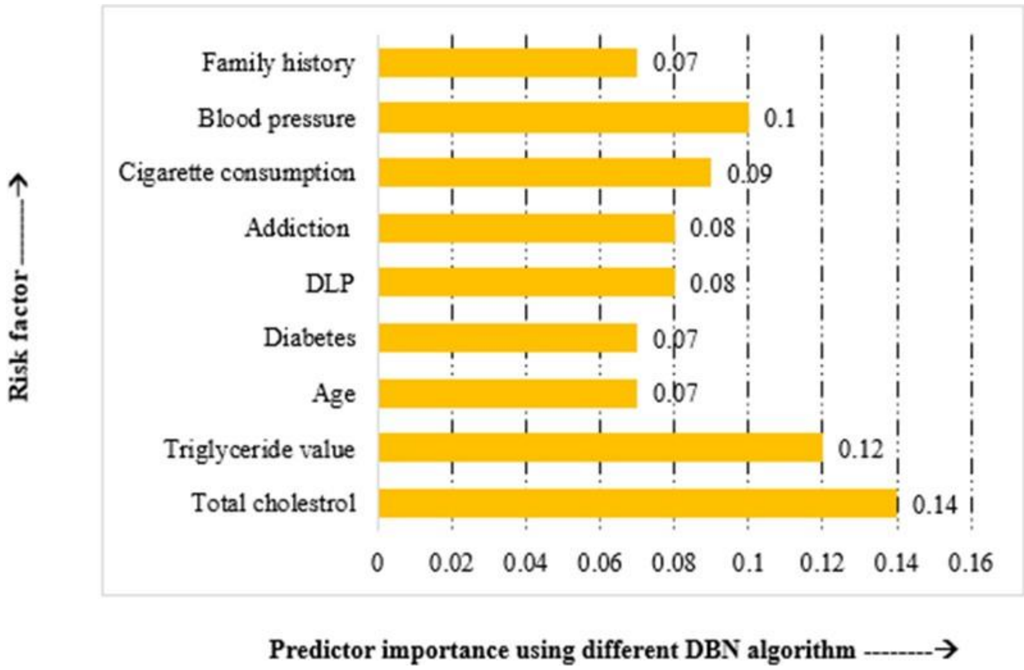
Rule No.	Rule Description	Rule accuracy
1	A heart attack cannot happen to someone who does not smoke.	68.6%
2	A person with a history of hypertension is protected against experiencing a MI.	68.6%
3	Even if there is no history of hypertension in the family, a person who smokes heavily has an increased risk of having a heart attack.	68.6%
4	A heart attack is not something that happens to someone who is not	64.3%

	addicted.	
5	A MI is likely to occur in a person with elevated triglyceride levels.	64.3%
6	A cardiac infarction would occur in a person younger than 55 with a history of substance abuse and high blood pressure.	64.3%
7	A heart attack is almost certain to occur in a smoker.	60%
8	A MI can occur in a person with a substance use disorder but normal blood pressure.	63.8%

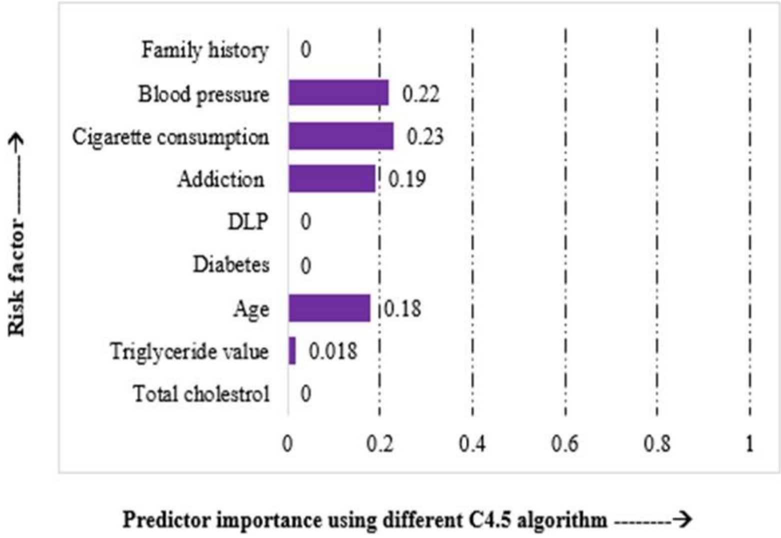
According to Figure 11a, the DBN algorithm found that elevated cholesterol (0.14), elevated triglycerides (0.12), elevated blood pressure can (0.10), nicotine (0.09), addictions (0.08), DLP (0.08), a high concentration of lipoprotein that is (HDL; 0.06) was discovered as variable increasing the chance of cardiovascular disease (CVD), whereas a diabetic problem (0.07), age (0.07), and familial risk of CVD (0.07) were the largest predictors of MI. As can be seen in Figure 11b, the C4.5 algorithm (0.18) identified many background characteristics as significant predictive of myocardial infarction (MI), including nicotine dependency, high blood pressure, age, and levels of triglycerides (0.23, 0.22, 0.19, and 0.10, respectively). Several risk factors for myocardial infarction were discovered by examination of RF algorithms. Smoking background (0.09), concentrations of LDL cholesterol (0.08), HDL cholesterol (0.08), hypertension (0.09), and BMI (0.09), lifespan (0.08), and diabetes (0.08) are all risk variables [Figure 11c]. Figure 11d depicts the results of the BN algorithm identifying the strongest predictors of MI as follows: cigarette use (0.33), LDL (0.17), addictive behaviors (0.14), cholesterol levels (0.09), the level of HDL (0.08), DLP (0.03), weight (0.02), lifespans (0.03), triglycerides (0.06), and hypertension to (0.02).

Table 6. The four algorithms utilized are averaged to reveal that "cigarette usage" is the most significant predictor of MI.

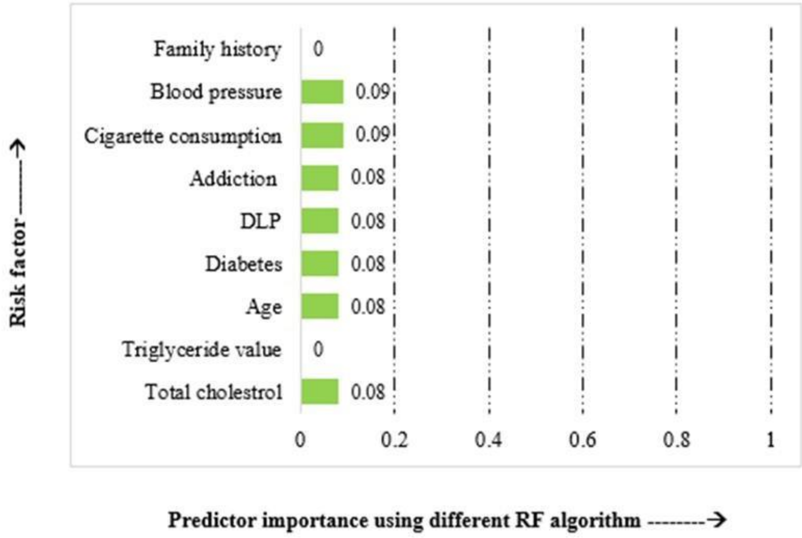
ID	Risk factor	Predictor importance using different algorithms				Average importance of predictors	Priority
		DBN	C4.5	RF	BN		
1	**Age**	0.07	0.18	0.08	0.03	0.09	4
2	**Gender**	0	0	0	0	0.00	11
3	**Family history**	0.07	0	0	0	0.02	10
4	**Body mass index**	0	0	0.09	0.02	0.03	9
5	**Cigarette consumption**	0.09	0.23	0.09	0.33	0.19	1
6	**Addiction**	0.08	0.19	0.08	0.14	0.12	2
7	**DLP**	0.08	0	0.08	0.03	0.07	5
8	**Triglyceride value**	0.12	0.018	0	0.06	0.05	7
9	**Low-density lipoprotein**	0	0	0.08	0.17	0.06	6
10	**High-density lipoprotein**	0	0	0.08	0.08	0.04	8
11	**Diabetes**	0.07	0	0.08	0	0.04	8
12	**Total cholesterol**	0.14	0	0.08	0.14	0.09	4
13	**Blood pressure**	0.10	0.22	0.09	0.02	0.11	3



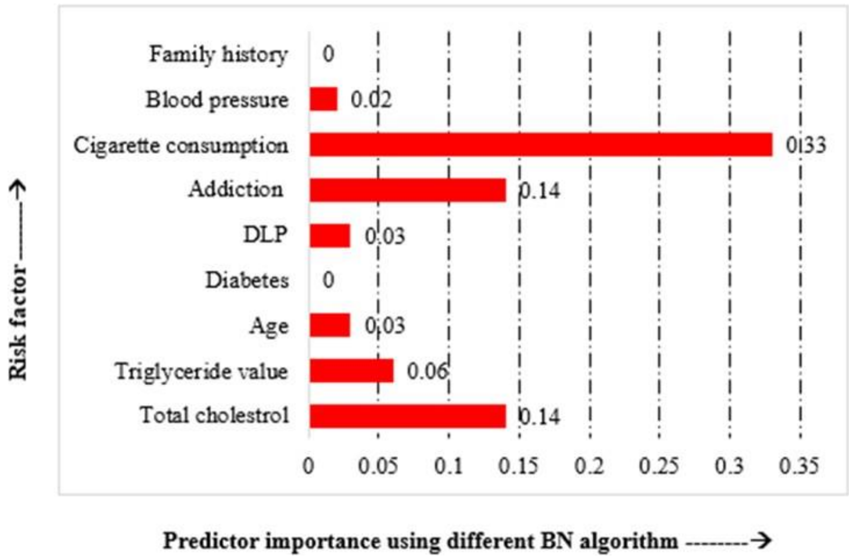
(a)



(b)



(c)



(d)

Figure 11. The significance of predictor variables a) using DBN algorithm b) using C4.5 algorithm c) using RF algorithm d) using BN algorithm

VI. Conclusion And Future Scope

When the heart's blood supply is abruptly interrupted or drastically reduced, it is called a Myocardial Infarction (MI), sometimes known as a heart attack. It might occur with no warning signs and go undiagnosed, or it may be a catastrophic occurrence that causes a fast decline in hemodynamic status and higher mortality. This study examines the complex domain of forecasting Myocardial Infarction (MI) by thoroughly investigating 40 distinct variables, encompassing aspects such as lifestyle, health indicators, and genetic factors. This paper aims to showcase the application of machine learning in predicting myocardial infarction based on gender. In this research author proposed a Deep belief network (DBN) approach that achieve an accuracy (89.32%), precision (84.04%), sensitivity (86.63%), and specificity (82.45%) in the prediction and diagnosis of the myocardial infarction. In this research the comparison of diverse machine learning algorithms, including Deep Belief Network (DBN), C4.5, Random Forest (RF), and Bayesian Network (BN), are provided and its significance in attaining remarkable accuracy rates. Author analyses that among these algorithms, DBN stands out with a leading accuracy rate of 89.32%.

This research makes a substantial contribution to the field by providing clinicians and researchers with a strong framework to comprehend and potentially prevent myocardial infarction, using individual patient profiles. This study establishes the groundwork for a more sophisticated comprehension of myocardial infarction (MI) prediction, utilizing the capabilities of artificial intelligence to navigate the complex terrain of cardiovascular well-being.

The prediction and diagnosis of myocardial infarction are further optimized by the ensemble classifier model. The continuous improvement of predictive models influences the application of this method and contributes to the improvement of prediction model accuracy. The application of ensemble classifiers in the diagnosis of myocardial infarction holds significant promise for the progression of medical diagnostics and the improvement of patient care. This study has the potential to significantly contribute to the field by applying ensemble classifiers and machine learning models to forecast gender-specific myocardial infarctions. As a result, diagnostic precision and patient outcomes could be revolutionized.

References

[1] L. Nagel, “The influence of the COVID-19 pandemic on the digital transformation of work,” Int. J. Soc. Soc. Policy, vol. 40, no. 9/10, pp. 861-875, 2020. doi:10.1108/IJSSP-07-2020-0323.

[2] D. Chumachenko et al., “Machine learning methods in predicting patients with suspected myocardial infarction based on short-time HRV data,” Sensors (Basel), vol. 22, no. 18, p. 7033, 2022. doi:10.3390/s22187033.

[3] I. R. Dégano, et al., “Twenty-five-year trends in myocardial infarction attack and mortality rates, and case-fatality, in six European populations,” Heart, vol. 101, no. 17, pp. 1413-1421, 2015. doi:10.1136/heartjnl-2014-307310.

[4] Y. S. Hamirani et al., “Effect of microvascular obstruction and intramyocardial hemorrhage by CMR on LV remodeling and outcomes after myocardial infarction: A systematic review and meta-analysis,” JACC Cardiovasc. Imaging, vol. 7, no. 9, pp. 940-952, 2014. doi:10.1016/j.jcmg.2014.06.012.

[5] N. Nandal et al., “‘Machine learning-based heart attack prediction: A symptomatic heart attack prediction method and exploratory analysis.’ F1000Research,” vol. 11, no. 1126, p. 1126, 2022.

[6] Y. Sandhy, “Prediction of heart diseases using support vector machine,” IJRASET, vol. 8, no. 2, 2020. doi:10.22214/ijraset.2020.2021 (ISSN, p. 2321, vol. 9653).

[7] R. J. Koene et al., “Shared risk factors in cardiovascular disease and cancer,” Circulation, vol. 133, no. 11, pp. 1104-1114, 2016. doi:10.1161/CIRCULATIONAHA.115.020406.

[8] M. Saeidi et al., “Gender differences in patients’ beliefs about biological, environmental, behavioral, and psychological risk factors in a cardiac rehabilitation program,” J. Cardio Thorac. Med., vol. 2, no. 4, pp. 215-220, 2014.

[9] M. A. Hogo, “A proposed gender-based approach for diagnosis of the coronary artery disease,” SN Appl. Sci., vol. 2, no. 6, pp. 1-17, 2020. doi:10.1007/s42452-020-2858-1.

[10] J. Soni et al., “Predictive data mining for medical diagnosis: An overview of heart disease prediction,” Int. J. Comput. Appl., vol. 17, no. 8, pp. 43-48, 2011. doi:10.5120/2237-2860.

[11] Valletta, John Joseph, Colin Torney, Michael Kings, Alex Thornton, and Joah Madden. "Applications of machine learning in animal behaviour studies." Animal Behaviour 124 (2017): 203-220

[12] Hakim, Md Azizul, Nusrat Jahan, Zannat Ara Zerin, and Amena Begum Farha. "Performance Evaluation and Comparison of Ensemble Based Bagging and Boosting Machine Learning Methods for Automated Early Prediction of Myocardial Infarction." In 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1-6. IEEE, 2021

[13] H. Mohan, Textbook of Pathophysiology, 5th ed. New Delhi: Jaypee Brothers Medical Publishers (p) ltd, 2005.

[14] A. K. Singh and R. K. Jat, “Myocardial infarction,” Himalayan J. Health Sci., pp. 16-32, 2021.

[15] H. Akbar et al., Acute ST-Elevation Myocardial Infarction, 2018.

[16] C. Gopalan and E. Kirk, Biology of Cardiovascular and Metabolic Diseases. Academic Press, 2022.

[17] Available at: https://my.clevelandclinic.org/health/diseases/16818-heart-attack-myocardial-infarction.

[18] Han, Chuang, Jiajia Sun, Yingnan Bian, Wenge Que, and Li Shi. "Automated Detection and Localization of Myocardial Infarction With Interpretability Analysis Based on Deep Learning." IEEE Transactions on Instrumentation and Measurement 72 (2023): 1-12

[19] Xiong, Ping, Simon Ming-Yuen Lee, and Ging Chan. "Deep learning for detecting and locating myocardial infarction by electrocardiogram: A literature review." Frontiers in cardiovascular medicine 9 (2022): 860032.

[20] B.-I. Bat-Erdene et al., "Deep learning-based prediction of heart failure rehospitalization during 6, 12, 24-month follow-ups in patients with acute myocardial infarction," Health Inform. J., vol. 28, no. 2, 14604582221101529, 2022. doi:10.1177/14604582221101529.

[21] Y. Xue et al., "Risk stratification of ST-segment elevation myocardial infarction (STEMI) patients using machine learning based on lipid profiles," Lipids Health Dis., vol. 20, no. 1, pp. 48, 2021. doi:10.1186/s12944-021-01475-z.

[22] Z. Bai et al., "Clinical feature-based machine learning model for 1-year mortality risk prediction of ST-segment elevation myocardial infarction in patients with hyperuricemia: A Retrospective Study," Comp. Math. Methods Med., vol. 2021, 7252280, 2021. doi:10.1155/2021/7252280.

[23] L. Li et al., "Gender-related difference in D-dimer level predicts in-hospital heart failure after primary PCI for ST-segment elevation myocardial infarction," Dis. Markers, vol. 2021, 7641138, 2021. doi:10.1155/2021/7641138.

[24] P. Liang et al., "Prediction of patients with heart failure after myocardial infarction" in IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2020, pp. 2009-2014. doi:10.1109/BIBM49941.2020.9313253.

[25] S. Kamalapurkar and S. G. G H, "Online portal for prediction of heart disease using machine learning ensemble method (PrHD-ML)" in IEEE Bangalore Humanitarian Technology Conference (B-HTC). IEEE, 2020, pp. 1-6. doi:10.1109/B-HTC50970.2020.9297918.

[26] S. Kayyum et al., "Data analysis on myocardial infarction with the help of machine learning algorithms considering distinctive or non-distinctive features" in International Conference on Computer Communication and Informatics (ICCCI). IEEE, 2020, pp. 1-7. doi:10.1109/ICCCI48352.2020.9104104.

[27] M. P. Than, et al., "Machine learning to predict the likelihood of acute myocardial infarction," Circulation, vol. 140, no. 11, pp. 899-909, 2019. doi:10.1161/CIRCULATIONAHA.119.041980.

[28] Liu, Wenhan, Qijun Huang, Sheng Chang, Hao Wang, and Jin He. "Multiple-feature-branch convolutional neural network for myocardial infarction diagnosis using electrocardiogram." Biomedical Signal Processing and Control 45 (2018): 22-32

[29] DİKER, Aykut, Zafer CÖMERT, and A. V. C. I. Engin. "A diagnostic model for identification of myocardial infarction from electrocardiography signals." Bitlis Eren University Journal of Science and Technology 7, no. 2 (2017): 132-139.

[30] M. Waqar et al., "An efficient SMOTE-based deep learning model for heart attack prediction," Sci. Program., vol. 2021, 1- 12, 2021. doi:10.1155/2021/6621622.

[31] G. E. Hinton et al., "A fast-learning algorithm for deep belief nets," Neural Comput., vol. 18, no. 7, pp. 1527-1554, 2006. doi:10.1162/neco.2006.18.7.1527.

[32] "Deep Learning 0.1 documentation: Deep Belief Networks". Available at: http://deeplearning.net/tutorial/DBN.html.

[33] W. Feng et al., 'A deep belief network-based machine learning system for risky host detection.' arXiv Preprint ArXiv:1801.00025, 2017.

[34] Chen et al., "Design and analysis for early warning of rotor UAV based on data driven DBN," Electronics, vol. 8, no. 11, p. 1350, 2019. doi:10.3390/electronics8111350.

[35] B. C. Latha, "C. and Carolin Jeeva, S," Inform. Med. Unlocked. Elsevier, vol. 16, p. 1 9, 2019. Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques."

[36] A. Bijalwan et al., "Botnet analysis using ensemble classifier," Perspect. Sci., vol. 8, pp. 502-504, 2016. doi:10.1016/j.pisc.2016.05.008.

[37] C. Aggarwal Charu and Z. C. Xiang, Mining Text Data. Springer New York Dordrecht Heidelberg London, vol. '12. Springer Science+Business Media, LLC, 2012.

[38] J. Ortigosa-Hernández et al., "Approaching sentiment analysis by using semi-supervised learning of multi-dimensional classifiers," Neurocomputing, vol. 92, pp. 98-115, 2012. doi:10.1016/j.neucom.2012.01.030.

[39] I. H. Witten and E. Frank, Data Mining: Practical Machine Learning Tools and Techniques. San Mateo, CA, USA: Morgan Kaufmann, 2016.

[40] S. Patil et al., "Predictive modeling for credit card fraud detection using data analytics," Procedia Comput. Sci., vol. 132, pp. 385-395, 2018. doi:10.1016/j.procs.2018.05.199.

[41] O. Mbaabu, 2020, “Introduction to the random forest in machine learning,” Berreskuratua-(e) Tik https://www. section. io/engineering-education/introduction-to-random-forest-in-machine-learning.).

[42] Noriega, Leonardo. "Multilayer perceptron tutorial." School of Computing. Staffordshire University 4, no. 5 (2005): 444.

[43] Popescu, Marius-Constantin, Valentina E. Balas, Liliana Perescu-Popescu, and Nikos Mastorakis. "Multilayer perceptron and neural networks." WSEAS Transactions on Circuits and Systems 8, no. 7 (2009): 579-588.

[44] D. Dolezel et al., “Examining predictors of myocardial infarction,” Int. J. Environ. Res. Public Health, vol. 18, no. 21, p. 11284, 2021. doi:10.3390/ijerph182111284.

[45] Centers for Disease Control and Prevention, “Know your risk for heart disease,” Heart Dis.. Available online: https://www. cdc. gov/heart disease/risk_factors. htm (accessed on 13 September 2021) (2019).

[46] A. Aaby et al., “Health literacy is associated with health behavior and self-reported health: A large population-based study in individuals with cardiovascular disease,” Eur. J. Prev. Cardiol., vol. 24, no. 17, pp. 1880-1888, 2017. doi:10.1177/2047487317729538.

[47] K. Gao et al., “The life-course impact of smoking on hypertension, myocardial infarction, and respiratory diseases,” Sci. Rep., vol. 7, no. 1, p. 4330, 2017. doi:10.1038/s41598-017-04552-5.

[48] M. Balling et al., “VLDL cholesterol accounts for one-half of the risk of myocardial infarction associated with apoB-containing lipoproteins,” J. Am. Coll. Cardiol., vol. 76, no. 23, pp. 2725-2735, 2020. doi:10.1016/j.jacc.2020.09.610.

[49] K. Malmberg and L. Rydén, “Myocardial infarction in patients with diabetes mellitus,” Eur. Heart J., vol. 9, no. 3, pp. 259 - 264, 1988. doi:10.1093/oxfordjournals.eurheartj.a062494