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

Acute Myocardial Infarction: Prediction and Patient Assessment through Different ML Techniques

Nudrat Fatima, Sifatullah Siddiqi

Submitted: 25/11/2023 **Revised**: 05/01/2024 **Accepted**: 15/01/2024

Abstract: Myocardial Infarction stands as a prevalent and severe ailment on a global scale. It ranks among the primary contributors to the world's highest mortality rates. Sometimes a Myocardial Infarction can show no symptoms at all. It is a disease that occurs when there is less supply of blood to the heart. In this research paper the main aim is to evaluate various techniques of Machine Learning to predict accurately the disease and the adverse effect of the risk factors. The different ML Techniques are applied on the dataset collected which includes 350 entries which includes some MI patients and some non-MI patients including men and women. The dataset is trained and then the Ensemble Classifiers are applied that increases prediction performance. The Ensemble Classifiers helps to improve gender specific prediction precision by merging classifier prediction.

Keywords: *Acute Myocardial Infarction, Machine Learning, Ensemble Classifiers, Classification methods, Logistic Regression, Random Forest Classifier, SVM*

1. Introduction

Myocardial Infarction, classified as a non-communicable disease, stands out as a leading factor contributing to the rising global mortality rates. In India, cardiovascular diseases (CVD) have reached epidemic proportions, with the incidence of heart diseases quadrupling in recent years. Lifestyle changes have had a profound impact on human health. Human health is negatively impacted by a number of risk factors, such as constrictive pericarditis, resting blood pressure (Trestbps), serum cholesterol (Chol), fasting blood sugar (Fbs), exercise-induced angina (Exang), maximum heart rate achieved (Thalach), ST depression induced by exercise relative to rest (Oldpeak), ST segment shift relative to exercise-induced increases in heart rate (Slope), coronary calcium scan (Ca), and thalassemia (Thal). One of the most common cardiovascular conditions in the trauma field, acute myocardial infarction (AMI), is a major factor in this case. Acute coronary syndrome is a general term for a number of illnesses affecting the heart and blood vessels that are included in the larger category of cardiovascular diseases (CVD).Acute Myocardial Infarction (AMI) can present as Non-ST Elevation Myocardial Infarction (NSTEMI), ST Elevation MI (STEMI), or unstable angina (UA). The term "acute coronary syndrome" (ACS) refers to a variety of diseases [2].

This research throws light over Myocardial Infarction (MI) that has been the most severe symptom of coronary heart disease. The most satiating attribute of this disease is the sudden death risk in the initial days and in later stages following myocardial infarction. The relevance of this research is substantially evaluating the mortality risk, that helps to improve the patient's health via different treatment strategies. Several research studies have been conducted to assess the findings and factors influencing the survival of individuals with myocardial infarction. The findings may nevertheless vary because of variations in the areas of socioeconomic geographical indicators [3,28]. The objective of this research is to develop an effective model for predicting and visually interpreting the risk of death following myocardial infarction. This involves highlighting key mortality variables and assessing the nature of their impact. This research helps to find the diseases like Hypertension and Coronary Artery Disease using DBN and Ensemble classifier. The research encompasses gathering data and employing Machine Learning Algorithms for the purpose of data classification. The data are collected for patients on gender basis suffering from myocardial infarction along with other symptoms.[4]

Assistant Professor Integral University, Lucknow
nudratfatima2@gmail.com, sifatullah.siddiqi@gmail.com

2. Related Works

Research on cardiovascular illness began to rely more on artificial intelligence and data mining in 2010 [29, 32, 34]. Scholars like "Kavitha, Ramakrishnan, and Manoj K. Singh" utilized training based on performance gradients for selecting data, and they suggested an evolutionary computation method to tackle the difficulties presented by adding new datasets [23, 35]. They found the ideal set of weights for performance by using genetic algorithms. Data mining became one of the most popular areas for analyzing and categorizing medical data in 2011 in order to make predictions and find patterns. In order to give patient-specific assessments or suggestions based on unique patient features, "Padmakumari K. N. Anooj" argued for the creation of a clinical decision support system [24]. The weighted fuzzy rules were generated from categorized datasets using an automated method. Decision tree classifiers were then used to generate decision tree rules, which helped create fuzzy rules for estimating the risk of cardiovascular diseases. Significant attempts were made in 2012 to forecast and diagnose heart disease using a range of data mining approaches. 'Minghao, Yongjun, Ho Sun, Jang, Keun' proposed an incremental decision tree induction approach that used an ensemble method to mine evolutionary diagnostic criteria for cardiac arrhythmia categorization. Their suggested model beat existing algorithms in terms of accuracy and efficiency [13,37]. Example data for diagnosing arrhythmia from the UCI Repository was used to test the performance of the proposed method. ECG signal data was evaluated for Heart Disease Classification, and the newly developed method was compared to current algorithms such as SVM, ANN, and Decision Tree.

In 2013, data mining approaches were predominantly employed in research within the Cardiac Heart Disease field. Jafar et al. published a dataset with 303 patients and 54 attributes that captured a variety of important aspects [14]. They also offered a feature development strategy to improve the dataset. Following that, information gain and confidence measures were used to assess the usefulness of characteristics related to coronary artery disease (CAD). To evaluate performance utilizing the expanded dataset, various approaches such as Nave Bayes, SMO, Bagging, and Artificial Neural Networks (ANN) were used. A unique algorithm was applied inside the feature development process to produce new features (LAD recognizer, LCX recognizer, RCA recognizer) aimed at assessing the existence of blockages in three arteries [43,38]. Any of these newly generated traits with elevated values indicates an increased chance of coronary artery disease (CAD). The analysis utilized Rapid Miner tools, and Information Gain and Gini Index indicators were calculated for various features [44,20].

A wide range of innovative methodologies, paired with data mining algorithms, were used in numerous studies in 2014.

'Hlaudi Daniel, Mosima's study attempted to use data mining techniques on patient datasets to forecast possible heart attacks, with an emphasis on identifying the model with the highest proportion of correct diagnosis predictions [45,43]. GA-KM and MPSO-KM data mining methods were used to cluster the heart disease dataset and estimate model accuracy [46]. Confusion matrices for these data mining methods were developed, assisting in the development of illness prognostic models based on specified features [47,41]. By 2015, Data Mining has established itself as an effective field in healthcare [45,59,42]. To assure reliable findings, certain data mining methods, such as Nave Bayes, were developed [48]. Randa et al. proposed a methodology that involves integrating the results of machine learning analyses conducted on diverse datasets with a specific focus on Coronary Artery Disease (CAD). This approach addresses challenges related to missing and inconsistent data that may arise during the data collection process. They applied C4.5 and Decision Tree algorithms to a unified dataset, leading to the generation of new decision trees [49,50]. The Cleveland Dataset was utilized in this analysis [31,40].

Big Data Analytics gained popularity in 2016 for tackling healthcare-related challenges. 'Purushottam, Kanak, and Richa' developed a framework to discover principles for predicting patients' risk levels based on supplied health characteristics [51]. To select the optimum subset of criteria for forecasting risk, a hill climbing algorithm was used.

In 2017, substantial breakthroughs in the healthcare sector, notably in the domain of cardiac disorders, demonstrated more effective and accurate methods to results. The RFRS algorithm was used by Xiao et al. to improve the efficiency and efficacy of categorizing heart disease diagnosis. They presented a hybrid classification technique to handle redundant and relevant characteristics, resulting in improved overall performance [52,58].

Throughout 2018, extensive research was conducted on the facets and associated risks of coronary artery disease, along with other ailments related to heart diseases.

'Kathleen, Julia' developed an improved deep neural network (DNN) learning method that would assist both patients and healthcare workers [53]. The goal was to increase the accuracy and reliability of identifying and prognosis in patients with cardiac disease [54]. This study made use of the Cleveland Clinic dataset. Diagnostic accuracy, F-score, specificity, precision, sensitivity , K-S tests, ROC curve, and were used to assess the approach's accuracy [52,55,59,60].Multiple integrated techniques were used in 2019 to improve illness detection and prediction accuracy. In the study by 'Senthilkumar, Chandrasegar, and Gautam,' a novel way to identifying key information using machine learning techniques was developed, eventually improving accuracy for forecasting cardiovascular disease. For prediction, the Hybrid Random Forest model was combined with a linear HRFLM model. HRFLM used an Artificial Neural Network with backpropagation [56] as input, containing 13 clinical characteristics. In the UCI dataset, weight ideas from the literature, such as Multiple Criteria Decision

Making (MCDM) applied to the Incremental Feature Selection (IFS) dataset, were employed [57]. The outcomes were examined and assessed in comparison to other standard approaches.

Author & Study	Contribution	Limitations
Noor Akhmad Setiawan, P.A. Venkatachalam and Ahmad Fadzil M.Hani(2009) Diagnosis of Coronary Artery Disease Using Artificial Intelligence Based Decision Support System	Created a fuzzy decision support system designed for diagnosing coronary artery disease based on evidence.[6]	The suggested rule selection method, based on RST, is capable of choosing only 27 rules.
K.Srinivas, Dr.G.Raghavendra Rao, Dr. A.Govardhan(2010) Analysis of Coronary Heart Disease and Prediction of Heart Attack in Coal Mining Regions Using Data Mining Techniques	Analysis of the Behavioral Risk Factor Surveillance System is done, survey to test. Whether self-reported cardiovascular disease rates are higher in Singareni coal mining regions in Andhra Pradesh state, India, compared to other regions. [7]	Model was evaluated based only on two measures. 1. accuracy 2. sensitivity
A. Sheik Abdullah, R.R.Rajalaxmi(2012) A Data mining Model for predicting the Coronary Heart Disease using Random Forest Classifier	A data mining model was crafted employing the Random Forest classifier, aiming to enhance prediction accuracy and explore diverse events associated with coronary heart disease (CHD). 5]	Primarily focused on forecasting the incidence of diverse events through the implementation of the Random Forest Classifier Algorithm.
Roohallah Alizadehsania, Jafar Habibia, Mohammad Javad Hosseinia, Hoda Mashayekhia, Reihane Boghrati a, Asma Ghandehariouna, Behdad Bahadorianb, Zahra Alizadeh Sanib(2013) A data mining approach for diagnosis of coronary artery disease.	Introducing a dataset named Z Alizadeh Sani, comprising 303 patients and 54 features, the research incorporates data mining approaches that leverage multiple impactful features. Additionally, a method for feature creation is suggested to enhance the richness of the dataset.[14]	Only the accuracy levels of the Data mining approaches are being analysed.
K. Cinetha, Dr. P. Uma Maheswari(2014) Decision Support System for Precluding Coronary Heart Disease (CHD)	Risk factors are classified based on the significant damages they cause, and data mining functionalities are employed to determine the level of risk through clustering analysis.15]	Categorization is exclusively applied to the risk factors.
Qingcai Chen , Haodi Li, Buzhou Tang , Xiaolong Wang, Xin Liu, Zengjian Liu , Shu Liu ,Weida	A challenge in clinical Natural Language Processing is presented, featuring a track focused on	Identification of Heart Disease Risk: 1) Risk factor exctraction.2) Time Attribute Identification.

Wang, Qiwen Deng , Suisong Zhu , Yangxin Chen , Jingfeng Wang.(2015) An automatic system to identify heart disease risk factors in clinical texts over time	identifying heart disease risk factors in clinical text over time.[16]	
Qurat-ul-ain Mastoi , Teh Ying Wah , Ram Gopal Raj, and Uzair Iqbal(2018) Automated Diagnosis of Coronary Artery Disease:A Review and Workflow	To pinpoint optimal methods and classifiers for Coronary Artery Disease (CAD) identification, two workflows are established, each based on distinct parameter sets. The automated diagnosis of Coronary Artery Disease involves instances A and B.10]	Despite advancements, further enhancement in CAD classification is essential, as ECG signals fail to furnish the necessary information.
Haleh Ayatollahi, Leila Gholamhosseini and Masoud Salehi(2019) Predicting coronary artery disease: a comparison between two data mining algorithms	Comparing the positive predictive value (PPV) of CAD using Artificial Neural Network (ANN) and SVM Algorithm and their distinction in terms of predicting CAD in some hospitals.[17]	The prediction was solely reliant on the outcomes of two algorithms.
Carlos Martin-Isla, Victor M. Campello, Cristian Izquierdo, Zahra Raisi-Estabragh, Bettina Baeßler, Steffen E. Petersen and Karim Lekadir(2020) Image-Based Cardiac Diagnosis With Machine Learning: A Review	Thorough review of recent works in this field and detailed presentation of Machine Learning methods.[1]	Implementation not available

Table 1: - Comprehensive Summary [3]

3. Methodology

The prediction of Myocardial Infarction with Machine Learning Technique used various steps as:

- Collection of data from various datasets like different hospitals and UCI Repository.
- Pre-Processing of Data by Importing Libraries.
- Training of Data is done by ML Algorithms
- Different ML Algorithms used.

High Risk Factor	Low Risk Factor
High Blood Pressure	*Smoking*
High LDL cholesterol levels	*Lack of Exercise*
Low HDL cholesterol levels	*Diet*
Obesity	*Age*
Thalassemia(Thal)	*Exercise Induced Angina*
Slope	*Gender*
Coronary Calcium Scan (Ca)	*Fasting Blood Sugar*

Table 2: Risk Factors Classification based on the Research

A. Collection of data

Data for the analysis and prediction of the disease is gathered from diverse sources, including various Indian Hospitals, the UCI Repository, and Kaggle. The hospitals that actively participated in the research are:

Impulse Hospital, Lucknow
Integral Institute of Medical Sciences, Lucknow
Smiling Hearts Cardiac Centre, Kanpur

Table 3: - List of Hospitals for Data Collection

B. Pre-Processing of Data

The second step after Data Collection is preprocessing step that is done with the help of Python Programming. Initially, we examined the dataset for any missing values and subsequently eliminated them.

C. Training of Data

The dataset collected is summarized and visualized after importing the necessary libraries. Following that, several algorithms are assessed to make predictions. The Categorial data are encoded. Finally, the dataset is divided, and feature scaling is implemented. Utilizing feature selection and feature extraction algorithms facilitates effective data reduction.

D. Different Machine Learning Algorithms

In this Research Paper the methods of Machine Learning Algorithms that are used are:

a) Deep Belief Network
b) Ensemble Classifiers

After the data training step, classification is done by numerous machine learning algorithms among which Support Vector Machine, Logistic Regression and Random Forest produced better results[5]. DBN and Ensemble classifier generates more accurate solutions than a single model, this has been the case in several machine learning competitions, where the winning solutions used these methods.[6,21]

- ≈ Ensemble classifier reduces the dispersion and spread of predictions and model performance.
- ≈ DBN learning algorithm scan the dataset and short the training time on GPU powered machine.

Deep Belief Network

One of the Deep Neural Network is DBN which is an unsupervised network like RBM's.

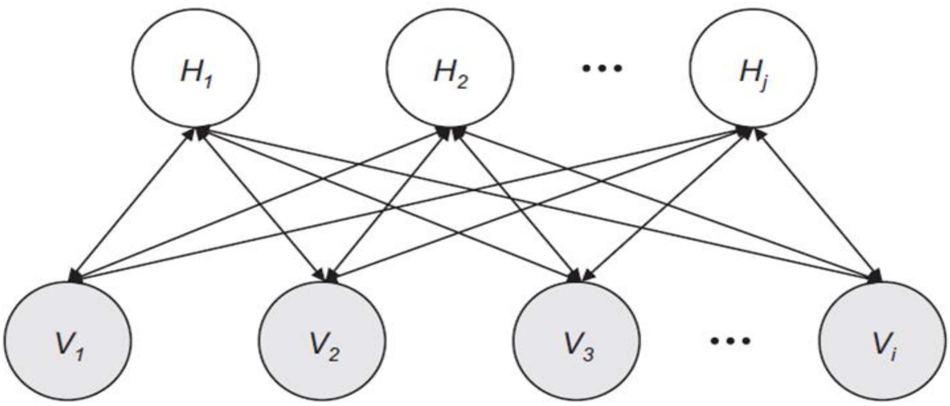
An RBM is made up of an input layer and a hidden layer, where the hidden layer uses connection weights W to detect characteristics in the data from the visible layer. Equations (2) and (3) may be used to explain this process mathematically, where v represents the input vector and h is the hidden vector [6,7,8]. As shown in Equation (4), the RBM energy-based model defines the joint distribution of the visible vector v and the hidden vector h, where E(v, h) represents the energy of a joint configuration of the hidden and visible units. Given a visible vector v, the hidden vector is stated as a conditional probability P(h|v). Likewise, the probability P(v|h) is calculated.

$$P(h/v) = \frac{P(h,v)}{Pv} \quad (2)$$

$$P(v/h) = \frac{P(v,h)}{P(h)} \quad (3)$$

$$P(h,v) = P(v,h) = \frac{1}{h}e^{-E(v,h)} \quad (4)$$

$$Z = \varepsilon_{v,h}e^{-E(v,h)} \quad (5)$$



The given figure 4 shows the process flow of Deep Belief Network [12]

Ensemble Classifier

This methodology is applied to boost the precision of classifiers. Employing this ensemble method proves effective in enhancing accuracy of various cardiac disease prediction systems.. The goal of integrating several classifiers is to achieve greater performance than a single classifier [22].

In the given dataset, training patterns are represented as {(x, t), (x, t), ..., (x, t)}, where each pattern is defined by an n-dimensional vector of continuous-valued features denoted as x.j = < x j1, x j2, ..., x jn >, and a class label tj with tj ∈ {class1, class2, ..., class Nclass}. A layer is identified as l, and the K clusters at layer l are denoted by Ω_(l,1), Ω_(l,2), ..., Ω_(l,k) where 1 ≤ l ≤ N layers. A structure in the training dataset is comparable to a vertex in n-dimensional Euclidean space.The primary goal of the clustering method is to categorize sets of data that exhibit mathematical similarity. For two patterns in the training set, represented as (xi, ti) and (xj, tj), their distance function d is defined in terms of their Euclidean distance. This distance function measures the mathematical dissimilarity between patterns, allowing for the categorization of data based on their similarity metrics.

$$\boldsymbol{d(x_i, x_j) = \sqrt{\sum_{k=1}^{n}(x_{ik} - x_{jk})^2}}$$

where "x_i =< x_{i1}, x_{i2}, ..., x_{in}>" and "x_j= < x $_{j1}$, x $_{j2}$, ..., x $_{jn}$>". Let a set of K-clusters $\Omega_{l,1}\Omega_{l,2} \ldots \Omega_{l,k}$ at layer l, the connected cluster centers Υ_l = {K_l,1, K_l,2, . . ., K_l, K} are coincidentally started, and the clustering algorithm seeks to minimize a optimal solution for each data set in the training dataset.

$j_{Li=\sum_{k=1}^{K}\sum_{\forall xj\in\Omega i,k} d(x_j,k_{i,k})}$ ….. 15

In the conclusion of the clustering procedure at layer l, each design (xi, ti) is allocated to a specific cluster Ω_(l,k), where 1 ≤ k ≤ K. This assignment results in clusters being classified as either atomic or non-atomic. For each cluster Ω_(l,k), a class-distributed vector is defined. This vector provides information about the distribution of classes within the cluster, delineating how different class labels are distributed among the patterns assigned to that particular cluster.

$$\Phi_{\Omega_{l,k}}(c_i) = \sum\nolimits_{\forall(x_j,t_j)\in\Omega_{l,k}} \Phi(t_j, c_j)$$

where

$\phi(t_j, c_i) = \{_{0}^{1} if\ t_j = c_j\ otherwise$ 17

and ci ∈ {$class_1$, $class_2$, ..., $classN_{class}$}. A cluster $\Omega_{l,k}$ is defined atomic if

$\frac{\max(\Phi\Omega_{l,k})}{\sum\forall c_i \Phi_{\Omega_{l,k}}(c_i)} = 1$ 18

A cluster that does not fulfil (5) contains patterns from several classes and is therefore considered non-atomic. An atomic cluster's $\Omega_{l,k}$ class designation is remembered as

$\beta_{l,k} = argmax\Phi_{\Omega_{l,k}}(c_i)$ 19

where ci ∈ {$class_1$, $class_2$, ..., $classN_{class}$}.

A neural network $\theta_{l,k}$ is trained on the system is consisted in each non-atomic cluster $\Omega_{l,k}$ to understand the decision boundaries. Classification of a test pattern **x** begins with the discovery of the suitable cluster at every layer. The distance between **x** and the center of each cluster $K_{l,k}$ is calculated using (1) and the suitable cluster at layer l is chosen as

$\Omega_{l,k} = argmin\ d(\boldsymbol{x}, k_{l,k})$ 14

The different Classifier that will be used for the prediction are: -

(i) Bayesian Network (BN)

The Naive Bayes (NB) classifier is based on the assumption that characteristics are independent of one another. In contrast, the opposing extreme hypothesis holds that all traits are interdependent. As a result of the Bayesian Network (BN) model, a directed acyclic graph is generated. The nodes in this graph represent random variables, while the edges depict conditional interdependence between these variables. The graph's structure illustrates the probabilistic linkages and dependencies between the dataset's distinct attributes. Models like this are called BN. Complete models for the variables and their connections are regarded as BNs. [25]. In this Research work a classification strategy is proposed based on multi-dimensional BN Classifiers. In this all the target values are grouped into one classification task focussing on the relationships between them.

(ii) Naïve Bayes Classifier (NBC)

One method of text categorization is allocated to a certain document d the class c^*

$= \arg max_c\ p(c|d).$

The development of the Naive Bayes (NB) classifier begins by applying Bayes' rule. This rule is a fundamental principle in probability theory that allows for the calculation of conditional probabilities by incorporating prior knowledge. In the context of the NB classifier, Bayes' rule is employed to estimate the probability of a particular class given observed features, taking into account the prior probability of that class and the likelihood of the observed features given that class.

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

[26]

where P(d) shows no role in choosing c^*. To approximate the term, P (d | c), NBC it by

supposing the $f_i's$ are provisionally individually given $d's$ class:

$$P_{NB}(c|d) := \frac{P(c)(\prod_{i=1}^{m} P(f_i|c)^{n_i(d)}}{P(d)} \quad [27]$$

The training procedure employs add-one smoothing to estimate the relative-frequency probabilities P(c) and P(f_i | c). When specific combinations of class (c) and feature (f_i) pairs are missing from the training data, this strategy is used to handle the problem of zero probability. By adding a pseudocount of one to all conceivable outcomes, add-one smoothing assures that even unseen combinations receive a non-zero probability, preventing probabilities from being wholly relied on observable data and avoiding the problem of zero probability for unseen occurrences.

(iii) Random Forest (RF)

An ensemble classifier, employed to enhance the accuracy of machine learning algorithms, is utilized for improved performance. Using RF techniques could assist in the identification of pertinent independent variables and empower the system to autonomously select its functionality.

Multiple studies have consistently confirmed the importance of choosing multiple options for each shrub in empirical investigations. This underscores the superiority of such an approach in respect to prediction accuracy.[36]

In Scikit-learn, Gini Significance is utilized to assess the importance of each node in a Decision Tree, particularly when there are only two child nodes.

$$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)} C_{right(j)}$$

[33]

Where, ni_j= the importance of node j

C_j = the impurity value of node j

w_j = weighted number of samples reaching node j

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

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

Where, ni_j represents the significance of node j, w_j is the weighted number of samples reaching node j, C_j 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.

(iv) C4.5

The foundational ID3 algorithm, a straightforward Decision Tree algorithm developed by Quinlan, serves as the basis for the C4.5 algorithm.[39] This algorithm divides trees based on the information gain ratio and takes input data, producing a Decision Tree as the output.

This method generates univariate trees, utilizing Decision Trees to represent classification rules. The tree splitting process ceases when it drops below a predefined threshold value, implementing pruning based on errors. This approach proves effective in handling numeric properties.[41,29]

(v) Multilayer Perceptron (MP)

The MP algorithm employs artificial neurons distributed across multiple layers, including hidden layers, to address challenges in binary classification[28,32].These biologically inspired algorithms utilize perceptrons as artificial neurons, each equipped with an activation function. The activation function maps each neuron's weighted inputs, resulting in the reduction of layers to two[20,21]. Perceptrons learn by adjusting the weights they are assigned in the process [66][67][68].

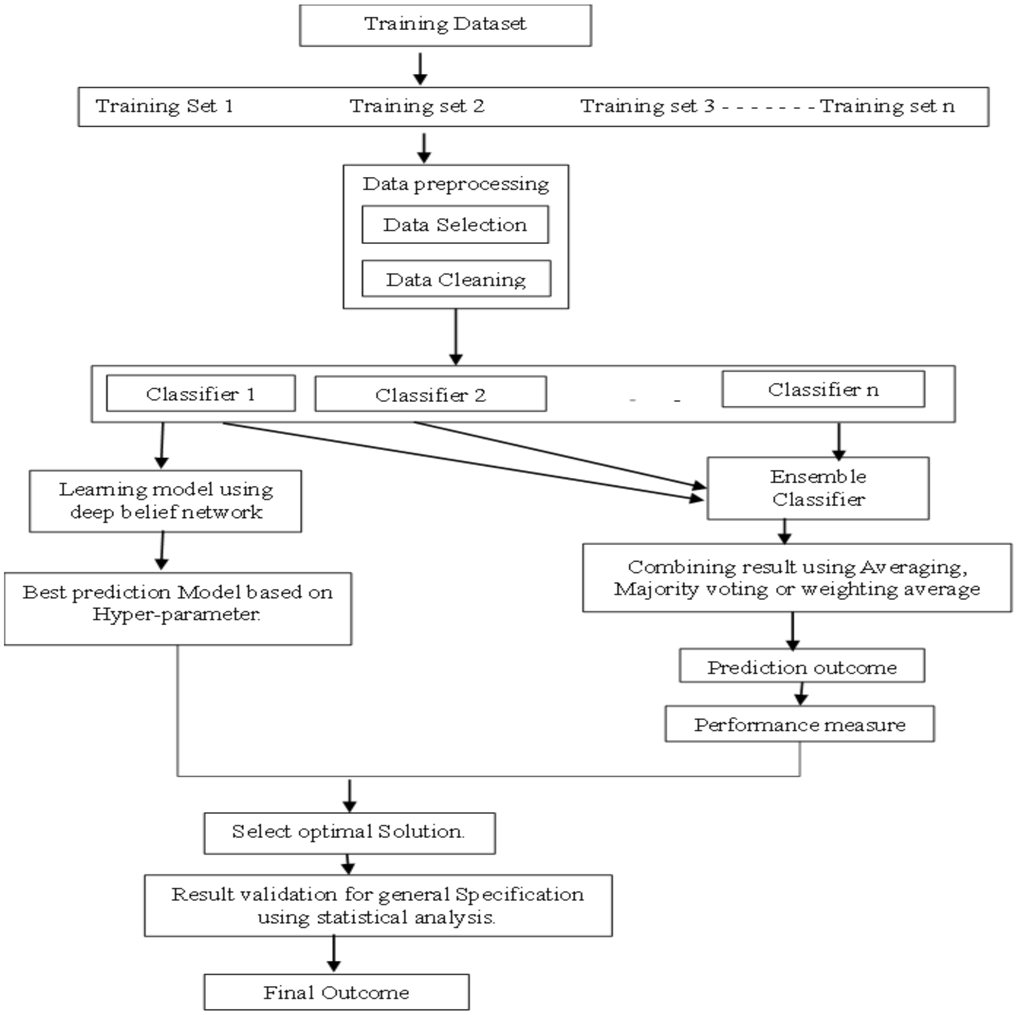


Figure 1: - Proposed Work

4. Proposed Parameters and Results

The effectiveness of the proposed technique was evaluated using DBN and ensemble classifiers (C4.5, RF, BN), achieving prediction accuracies of 89.32%, 87.53%, 81.29%, and 76.59%, respectively, as detailed in Table 4—marking them as the most successful modeling strategies.[30] The rules generated by the C5 algorithm are provided in Table 5. Figure 10 illustrates a graphical representation of the proposed methods, considering accuracy, precision, sensitivity, and specificity [61][62].

The evaluation of the suggested technique employs a set of performance metrics for assessment.

- $Total\_Images$ The sum amount of examined pictures.
- T_P (True positive): Detected the altered images without error.
- T_N (True negative): Validated as authentic on visual inspection.
- F_N (False negative): falsely recognized manipulated images or manipulated images that were mistakenly thought to be genuine.
- F_P (False positive): images that have been mistakenly recognized as being genuine or as being manipulated.

The formulas for these are as follows:

$$Accuracy = \frac{T_P+T_N}{T_{Total\_Images}} \times 100 \quad [7]$$

$$Precission = \frac{T_P}{T_P+F_P} \quad [8]$$

$$Recall = \left(\frac{T_P}{T_P+F_N}\right) \quad [9]$$

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%

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

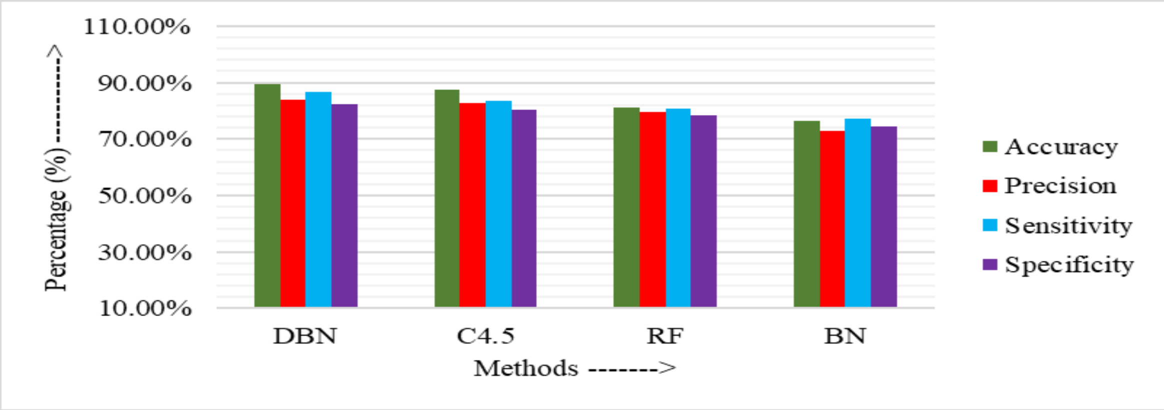


Fig 5: - A visual representation of the proposed method is depicted in the graph.

The DBN algorithm highlighted specific variables, such as elevated cholesterol (0.14), elevated triglycerides (0.12), elevated blood pressure (0.10), nicotine (0.09), addictions (0.08), and DLP (0.08), along with a high concentration of lipoprotein (HDL; 0.06), as contributors to increased cardiovascular disease (CVD) risk. Conversely, a diabetic condition (0.07), age (0.07), and familial risk of CVD (0.07) emerged as the predominant predictors of myocardial infarction (MI). In contrast, the C4.5 algorithm (0.18), illustrated in Figure identified various background [57]characteristics, including nicotine dependency, high blood pressure, age, and triglyceride levels (0.23, 0.22, 0.19, and 0.10, respectively), as significant predictors of MI.[31]The RF algorithms uncovered several risk factors for MI, with smoking history (0.09), hypertension (0.09), BMI (0.09), LDL cholesterol levels (0.08), HDL cholesterol levels (0.08), lifespan (0.08), and diabetes (0.08) identified as risk variables (Figure 11c). Figure 11d presents the BN algorithm results, identifying cigarette use (0.33), LDL (0.17), addictive behaviors (0.14), cholesterol levels (0.09), HDL level (0.08), triglycerides (0.06), DLP (0.03), lifespan (0.03), weight (0.02), and hypertension (0.02) as the strongest predictors of MI [63][64][65].

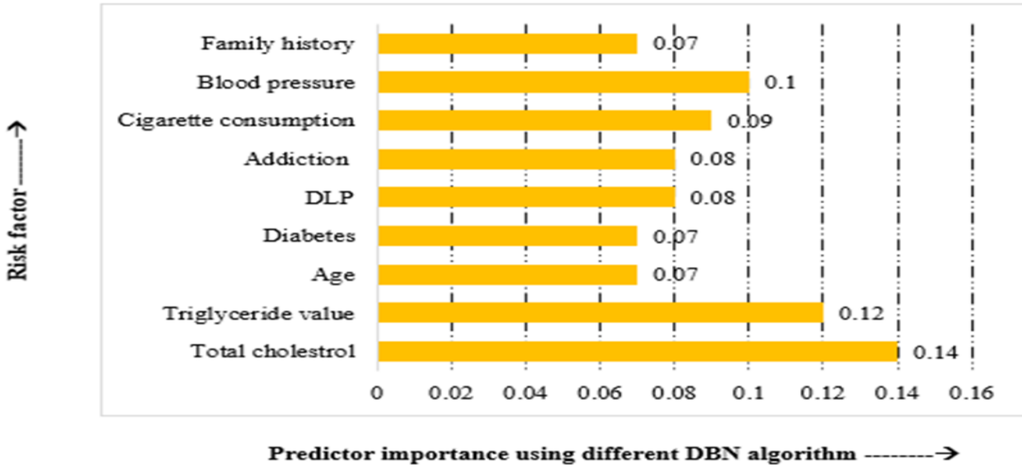


Fig 6

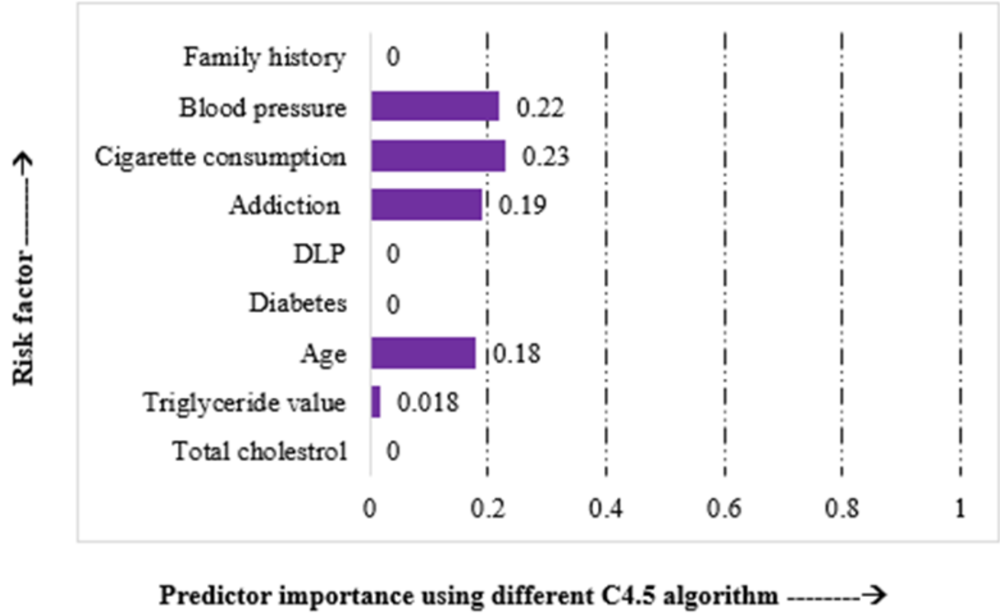


Fig 7

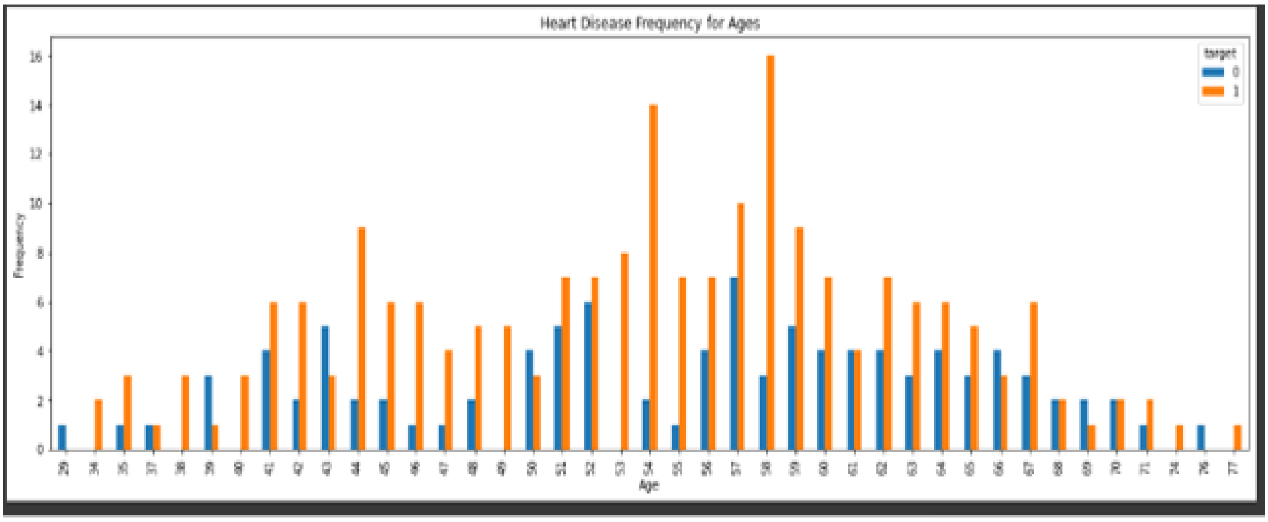


Fig 8

The figure illustrates the frequency of heart disease occurrence in both test and train data across different genders, as determined by the Decision Tree.

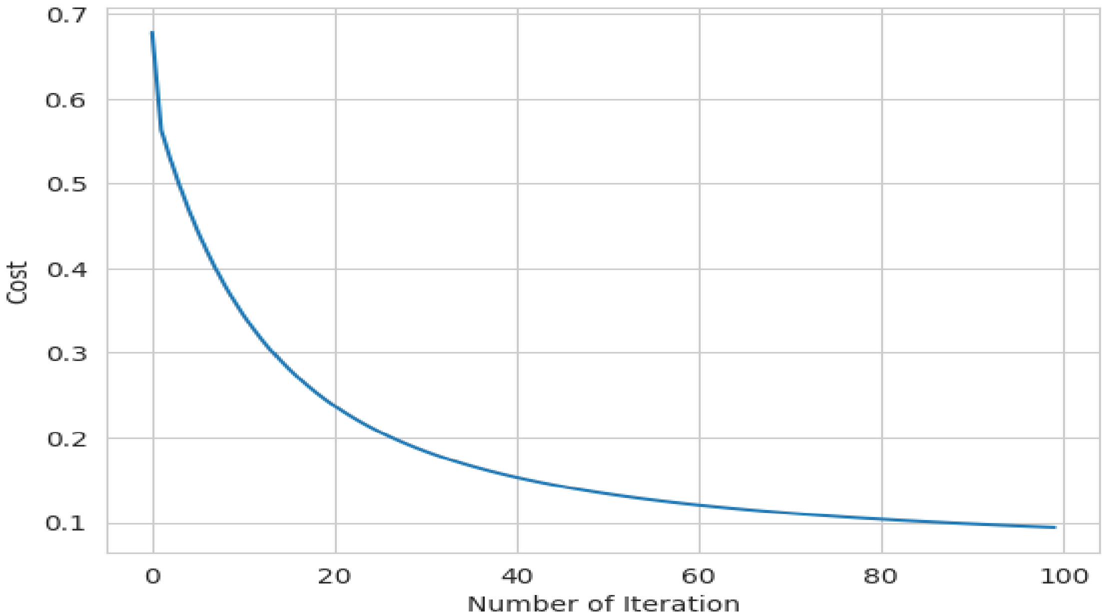
```
Classification Report for SVM:
              precision    recall  f1-score   support

           0       1.00      0.85      0.92        20
           1       0.93      1.00      0.96        41

    accuracy                           0.95        61
   macro avg       0.97      0.93      0.94        61
weighted avg       0.95      0.95      0.95        61
```

```
Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.85      0.92        20
           1       0.93      1.00      0.96        41

    accuracy                           0.95        61
   macro avg       0.97      0.93      0.94        61
weighted avg       0.95      0.95      0.95        61
```

```
Classification Report for Naive Bayes:
              precision    recall  f1-score   support

           0       1.00      0.85      0.92        20
           1       0.93      1.00      0.96        41

    accuracy                           0.95        61
   macro avg       0.97      0.93      0.94        61
weighted avg       0.95      0.95      0.95        61
```

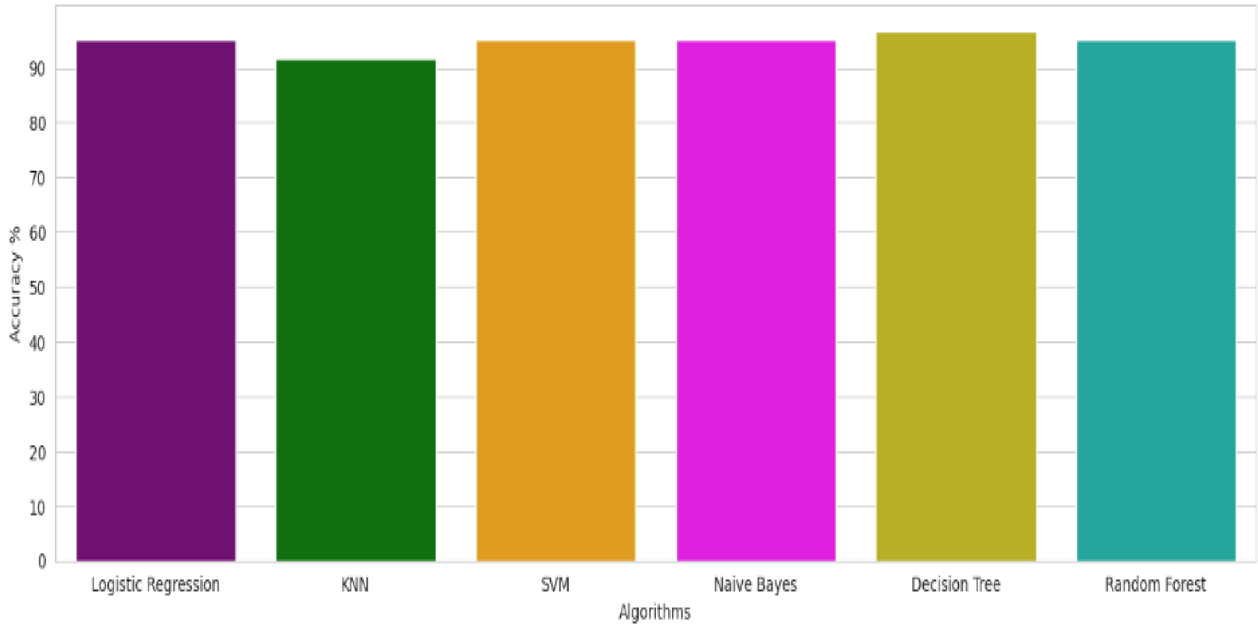
```
Classification Report for Decision Tree:
              precision    recall  f1-score   support

           0       1.00      0.90      0.95        20
           1       0.95      1.00      0.98        41

    accuracy                           0.97        61
   macro avg       0.98      0.95      0.96        61
weighted avg       0.97      0.97      0.97        61
```

```
iteration: 100
cost: 0.0942556751040842
```




```
Manuel Test Accuracy: 95.08%
```

Graph A



Graph B

5. Conclusion and Future Scope

Myocardial Infarction (MI), commonly known as a heart attack, occurs when the blood supply to the heart is suddenly cut off or significantly reduced. This event may unfold without warning signs, leading to undiagnosed cases, or it can manifest as a catastrophic incident resulting in a rapid decline in hemodynamic status and increased mortality. Coronary Artery Disease (CAD), the leading cause of mortality in the United States, underlies most instances of MI. The myocardium experiences oxygen deprivation with coronary artery occlusion, potentially leading to myocardial cell death and necrosis due to prolonged hypoxia.

Diagnosing Cardiovascular Disease (CVD) involves methods such as electrocardiography, ultrasonography, angiography, blood tests, and other diagnostic tools. However, these procedures are time-consuming and expensive, requiring multiple tests. This paper aims to showcase the utilization of Machine Learning (ML) for predicting MI based on gender. Recently developed ML-based CVD prediction techniques seek to enhance existing diagnostic methods. The experiment's results reveal that the developed Deep Belief Network (DBN) approach demonstrates notable accuracy (89.32%), precision (84.04%), sensitivity (86.63%), and specificity (82.45%). Future research should focus on advancing model development and practical implementation.

Acknowledgement:

In accordance with the guidelines of university doctoral studies and research, we are pleased to acknowledge the assignment of Manuscript Communication Number [IU/R&D/2024-MCN0002337] to this article. This identifier facilitates efficient communication and tracking of our research throughout the publication process. We extend our gratitude to all those who have contributed to the development of this work.

References:

[1] Dégano IR, Salomaa V, Veronesi G, Ferriéres J, Kirchberger I, Laks T, et al. Twenty-five-year trends in myocardial infarction attack and mortality rates, and case-fatality, in six European populations. Heart. (2015) 101:1413–21. doi: 10.1136/heartjnl-2014-307310.

[2] Singh A, Museedi AS, Grossman SA: Acute Coronary Syndrome. StatPearls [Internet], Treasure Island (FL): StatPearls Publishing; 2022.

[3] Nudrat Fatima, Mohammadi Akheela Khanum, An in Depth Review of the various Intelligent Approaches applied In the Detection and Prediction of Coronary Artery Disease (2021). Journal of Huazhong University of Science and Technology. (1671-4512) Vol 50.

[4] Carlos Martin-Isla, Victor M. Campello, Cristian Izquierdo, Zahra Raisi-Estabragh, Bettina Baeßler , Steffen E. Petersen and Karim Lekadir "Image-Based Cardiac Diagnosis With Machine Learning: A Review" (2020) Frontiers in Cardiovascular Medicine published: 24 January 2020, Volume 7, Article 1. doi: 10.3389/fcvm.2020.00001

[5] Minghao Piao1, Yongjun Piao1, Ho Sun Shon1, Jang-Whan Bae2, and Keun Ho Ryu1, "Evolutional Diagnostic Rules Mining for Heart Disease Classification Using ECG Signal Data"(2012) ©

Springer-Verlag Berlin Heidelberg 2012 Advances in Control and Communication, LNEE 137,pp. 673–680.

[6] Nagaraj M. Lutimath, Chethan C, Basavaraj S Pol " Prediction of Heart Disease using Machine Learning" (2019) presented at International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-2S10, September 2019.

[7] Prerna Jain , Amandeep Kaur " Big Data Analysis for Prediction of Coronary Artery Disease" (2018) presented at 4th International Conference on Computing Sciences.

[8] Qurat-ul-ain Mastoi , Teh Ying Wah , Ram Gopal Raj, and Uzair Iqbal "Automated Diagnosis of Coronary Artery Disease:A Review and Workflow" (2018). Cardiology Research and Practice Volume 2018, Article ID 2016282, 9 pages doi.org/10.1155/2018/2016282.

[9] Moloud Abdar, Sharareh R. Niakan Kalhori, Tole Sutikno, Imam Much Ibnu Subroto, Goli Arji "Comparing Performance of Data Mining Algorithms in Prediction Heart Diseases"(2015) International Journal of Electrical and Computer Engineering (IJECE) Vol. 5, No. 6, December 2015, pp. 1569~1576 ISSN: 2088-8708

[10] P.K. Anooj "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules" (2012) Journal of King Saud University – Computer and Information Sciences (2012) 24, 27–40

[11] Divya Krishnani, Anjali Kumari, Akash Dewangan, Aditya Singh, Nenavath Srinivas Naik "Prediction of Coronary Heart Disease using Supervised Machine Learning Algorithms"(2019) 2019 IEEE Region 10 Conference (TENCON 2019) 978-1-7281-1895-6/19/$31.00_c 2019 IEEE.

[12] Khan, M. U., & Ahamad, F. (2022). Multimodal Data Analysis and Machine Learning Techniques: A Comparison and Review. NeuroQuantology, 20(22), Article NQ10423. https://doi.org/10.48047/nq.2022.20.22.NQ10423D.

[13] 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.

[14] K. Cinetha, Dr. P. Uma Maheswari "Decision Support System for Precluding Coronary Heart Disease(CHD)" (2014) International Journal of Computer Science and Mobile Computing, Vol.3 Issue.2, February- 2014, pg. 34-38 ISSN 2320–088X.

[15] Qingcai Chen , Haodi Li, Buzhou Tang , Xiaolong Wang, Xin Liu, Zengjian Liu , Shu Liu ,Wang, Qiwen Deng , Suisong Zhu , Yangxin Chen , Jingfeng Wang. "An automatic system to identify heart disease risk factors in clinical texts over time." (2015) J Biomed Inform (2015), http://dx.doi.org/10.1016/j.jbi.2015.09.002

[16] Haleh Ayatollahi, Leila Gholamhosseini and Masoud Salehi "Predicting coronary artery disease: a comparison between two data mining algorithm" (2019) BMC Public Health (2019) 19:448 https://doi.org/10.1186/s12889-019-6721-5

[17] Md Saiful Islam 1, Md Mahmudul Hasan 1, Xiaoyi Wang 1, Hayley D. Germack 1,2,3 and Md Noor-E-Alam "A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining" (2018) Healthcare 2018, 6, 54 19. Anju Grewal, Hanish Kataria, and Ira Dhawan "Literature search for research planning and identification of research problem"(2016) PMC International journal of Anesthesia. Indian J Anaesth. 2016 Sep; 60(9): 635–639. doi: 10.4103/0019-5049.190618

[18] Mohammad Waqialla, Muhammad Imran Razzak "An ontology-based framework aiming to support cardiac rehabilitation program"(2016) 20th International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2016, 5-7 September 2016, York, United Kingdom Procedia Computer Science 96 (2016) 23 – 32, Elsevier

[19] Padmakumari K. N. Anooj " Clinical decision support system: risk level prediction of heart disease using weighted fuzzy rules and decision tree rules " (2011) Central European Journal of Computer Science 1(4) • 2011 • 482-498 DOI: 10.2478/s13537-011-0032-y.

[20] Manish Madhava Tripathi, Dr S P Tripathi, "Image Watermarking Based On Wavelet-Based Counterlet Packet Transformation With Best Tree", Internation Journal of Computer Application, 0975- 8887, Vol 13, no 4, feb 10.

[21] Jyoti Soni, Uzma Ansari, Dipesh Sharma "Intelligent and Effective Heart Disease

[22] Prediction System using Weighted Associative Classifiers" (2011) International Journal on Computer Science and Engineering (IJCSE).

[23] Sivagowry .S, Dr. Durairaj. M And Persia.A "An Empirical Study On Applying Data Mining

Techniques For The Analysis And Prediction Of Heart Disease" (2013) International Conference On Information Communication And Embedded Systems Icices 2013 978-1-4673-5788-3/13/$31.00©2013IEEE.

[24] Gayathri. P and N. Jaisankar "Comprehensive Study of Heart Disease Diagnosis Using Data Mining and Soft Computing Techniques"(2013) International Journal of Engineering and Technology (IJET) ISSN : 0975-4024 Vol 5 No 3 Jun-Jul 2013

[25] Hlaudi Daniel Masethe, Mosima Anna Masethe " Prediction of Heart Disease using Classification Algorithms" (2014) Proceedings of the World Congress on Engineering and Computer Science 2014 Vol II WCECS 2014, 22-24 October, 2014, San Francisco, USA ISBN: 978-988-19253-7-4 ISSN: 2078-0958 (Print); ISSN: 2078-0966

[26] Mohammad Subhi Al-batah "Testing the Probability of Heart Disease Using Classification and Regression Tree Model" (2014) Annual Research & Review in Biology, 4(11): 1713-1725, 2014.

[27] V. Krishnaiah, G. Narsimha, and N. Subhash Chandra "Heart Disease Prediction System Using Data Mining Technique by Fuzzy K-NN Approach" (2015) © Springer International Publishing Switzerland 2015,Emerging ICT for Bridging the Future – Volume 1, Advances in Intelligent Systems and Computing 337, DOI: 10.1007/978-3-319-13728-5_42

[28] Randa El-Bialy , Mostafa A. Salamay, Omar H. Karam and M.Essam Khalifa "Feature Analysis of Coronary Artery Heart Disease Data Sets"(2015) Procedia Computer Science 65 (2015) 459 – 4681877-0509 © 2015 The Authors. Published by Elsevier B.V. doi: 10.1016/j.procs.2015.09.132

[29] Salma Banu N.K, Suma Swamy "Prediction of Heart Disease at early stage using Data Mining and Big Data Analytics: A Survey" (2016) 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT)

[30] Theresa Princy. R, J. Thomas "Human Heart Disease Prediction System using Data Mining Techniques"(2016) 2016 International Conference on Circuit, Power and Computing Technologies [ICCPCT]

[31] Purushottam , Prof. (Dr.) Kanak Saxena, Richa Sharma "Efficient Heart Disease Prediction System" (2016) Procedia Computer Science 85 (2016) 962 – 969, Elsevier

[32] Manish Madhava Tripathi, Saurabh Pandey, "Diagnosis of Diabetes using Artificial Intelligence Techniques by using Bio Medical Signal Data", International Journal of Research and Development in Applied Science and Engineering (IJRDASE) ISSN-2454-6844, Volume 13, Issue 2, May 2017.

[33] Xiao Liu, Xiaoli Wang,Qiang Su,Mo Zhang, Yanhong Zhu, Qiugen Wang, and QianWang"A Hybrid Classification System for Heart Disease Diagnosis Based on the RFRS Method"(2017) Computational and Mathematical Methods in Medicine Volume 2017, Article ID 8272091, 11 pages https://doi.org/10.1155/2017/8272091

[34] Animesh Hazra, Subrata Kumar Mandal, Amit Gupta, Arkomita Mukherjee and Asmita Mukherjee "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review" (2017) Advances in Computational Sciences and Technology ISSN 0973-6107 Volume 10, Number 7 (2017) pp. 2137-2159 © Research India Publications

[35] Mujtaba Ashraf Qureshi, Dr. Irshad Ahmad Mir "Comparative Study of Existing Techniques for Heart Diseases Prediction Using Data Mining Approach" (2017) Asian Journal of Computer Science and Information Technology 7: 3 July (2017) 50 – 56.

[36] Kathleen H. Miao, Julia H. Miao "Coronary Heart Disease Diagnosis using Deep Neural Networks" (2018) (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9, No. 10, 2018.

[37] D. J. S. Sako & J. Palimote "A Medical Document Classification System for Heart Disease Diagnosis Using Naïve Bayesian Classifier" (2018) International Journal of Applied Science and Mathematical Theory ISSN 2489-009X Vol. 4 No. 1 2018

[38] Debabrata Swain, Santosh Kumar Pani, Debabala Swain "An Efficient System for the Prediction of Coronary Artery Disease using Dense Neural Network with Hyper Parameter Tuning" (2019) International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-6S, April 2019.

[39] Ravindhar NV, Anand, Hariharan Shanmugasundaram, Ragavendran, Godfrey Winster "Intelligent Diagnosis of Cardiac Disease Prediction using Machine Learning"(2019) International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-11, September 2019.

[40] Avinash Golande, Pavan Kumar T "Heart Disease Prediction Using Effective Machine Learning Techniques" (2019) International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-1S4, June 2019

[41] R.O. Bonow, D.L. Mann, D.P. Zipes, P. Libby, Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine, 9th ed., Saunders, New York, 2012.

[42] Mohammad Suaib, Dr. M. Akheela Khanum, "Web Page Personalization Techniques in the Purview of Page Ranking Methodology using Artificial Intelligence" in International Journal of Computer and Information Technology, Volume 8, Issue 6, ISSN: 2279-0764, November 2019.

[43] P.N. Tan, M. Steinbach, V. Kumar, Introduction to Data Mining, Pearson Addison Wesley, Boston, MA, 2006.

[44] J. Liu, Y.-T. HSU, and C.-L. Hung, "Development of Evolutionary Data Mining Algorithms and their Applications to Cardiac Disease Diagnosis," in WCCI 2012 IEEE World Congress on Computational Intelligence, 2012, pp. 10–15.

[45] https://www.analyticsvidhya.com/blog/2020/04/confusion-matrix-machine-learning/

[46] Meena Preethi. B, Darshna. R, Sruthi. R "Prediction of Coronary Artery Disease Using Text Mining" (2018) International Journal of Trend in Scientific Research and Development (IJTSRD) ISSN No: 2456 - 6470 | Volume - 2 | Issue – 6 | Sep – Oct 2018.

[47] Sabyasachi Dash, Sushil Kumar Shakyawar, Mohit Sharma and Sandeep Kaushik "Big data in healthcare: management,analysis and future prospects"(2019) Dash et al. J Big Data (2019) 6:54 https://doi.org/10.1186/s40537-019-0217-0

[48] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, The MIT Press, 2016.

[49] E. G. Nabel and E. Braunwald, "A tale of coronary artery disease and myocardial infarction," New England Journal of Medicine, Vol. 366, pp. 54-63, 2012.

[50] Faiz, M., & Daniel, A. K. (2023). A hybrid WSN based two-stage model for data collection and forecasting water consumption in metropolitan areas. International Journal of Nanotechnology, 20(5-10), 851-879.

[51] J. H. Miao and K. H. Miao, "Cardiotocographic diagnosis of fetal health based on multiclass morphologic pattern predictions using deep learning classification," International Journal of Advanced Computer Science and Applications, Vol. 9, No. 5, pp. 1-11, 2018.

[52] Senthilkumar Mohan, Chandrasegar Thirumalai,And Gautam Srivastava (Member, Ieee) "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques" (2019) Digital Object Identifier 10.1109/ACCESS.2019.2923707 IEEE Access

[53] F. Dammak, L. Baccour, and A. M. Alimi, ``The impact of criterion weights techniques in TOPSIS method of multi-criteria decision making in crisp and intuitionistic fuzzy domains," in Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE), vol. 9, Aug. 2015, pp. 1_8.

[54] A.-E. Hassanien, "Rough set approach for attribute reductionand rule generation: a case of patients with suspected breast cancer," Journal of the American Society for Information Science and Technology, vol. 55, no. 11, pp. 954–962, 2004.

[55] J. H. Miao, K. H. Miao, and G. J. Miao, "Breast cancer biopsy predictions based on mammographic diagnosis using Support Vector Machine learning," Multidisciplinary Journals in Science and Technology, Journal of Selected Areas in Bioinformatics, Vol. 5, No. 4, pp. 1–9, 2015

[56] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation.," A. Sattar and B. Kang (eds), Advances in Artificial Intelligence, Lecture Notes in Computer Science, Vol. 4304, Berlin, Heidelberg, 2006.

[57] T. B. Arnold and J. W. Emerson, "Nonparametric goodness-of-fit tests for discrete null distributions," The R Journal, Vol. 3/2, pp. 34-39, 2011.

[58] Yang Yan, Jia-Wen Zhang, Guang-Yao Zang, and Jun Pu "The primary use of artificial intelligence in cardiovascular diseases: what kind of potential role does artificial intelligence play in future medicine?" (2019) J Geriatr Cardiol. 2019 Aug; 16(8): 585–591. doi: 10.11909/j.issn.1671-5411.2019.08.010

[59] A.H. Nandhu Kishore "Automatic stenosis grading system for diagnosing coronary artery disease using coronary angiogram"(2019) Int. J. Biomedical Engineering and Technology, Vol. 31, No. 3, 2019

[60] J. A. Patel and P. Sharma, "Big data for Better Health Planning," (2014)Adv. Eng. Technol. Res. (ICAETR), 2014 Int. Conf. IEEE., pp. 0–4,2014.

[61] A. Srivastava and S. Ahmad, "Bio-Computing Based Algorithms for Cloud Security: A Critical Review," 2022 IEEE World Conference on Applied Intelligence and Computing (AIC), Sonbhadra,

India, 2022, pp. 894-900, doi: 10.1109/AIC55036.2022.9848965.

[62] Narayan, Vipul, et al. "7 Extracting business methodology: using artificial intelligence-based method." Semantic Intelligent Computing and Applications 16 (2023): 123

[63] Narayan, Vipul, et al. "A Comprehensive Review of Various Approach for Medical Image Segmentation and Disease Prediction." Wireless Personal Communications 132.3 (2023): 1819-1848.

[64] Mall, Pawan Kumar, et al. "Rank Based Two Stage Semi-Supervised Deep Learning Model for X-Ray Images Classification: AN APPROACH TOWARD TAGGING UNLABELED MEDICAL DATASET." Journal of Scientific & Industrial Research (JSIR) 82.08 (2023): 818-830.

[65] Narayan, Vipul, et al. "Severity of Lumpy Disease detection based on Deep Learning Technique." 2023 International Conference on Disruptive Technologies (ICDT). IEEE, 2023.

[66] Saxena, Aditya, et al. "Comparative Analysis Of AI Regression And Classification Models For Predicting House Damages İn Nepal: Proposed Architectures And Techniques." Journal of Pharmaceutical Negative Results (2022): 6203-6215.

[67] Kumar, Vaibhav, et al. "A Machine Learning Approach For Predicting Onset And Progression"“Towards Early Detection Of Chronic Diseases “." Journal of Pharmaceutical Negative Results (2022): 6195-6202.

[68] Chaturvedi, Pooja, Ajai Kumar Daniel, and Vipul Narayan. "Coverage Prediction for Target Coverage in WSN Using Machine Learning Approaches." (2021).

[69] Chaturvedi, Pooja, A. K. Daniel, and Vipul Narayan. "A Novel Heuristic for Maximizing Lifetime of Target Coverage in Wireless Sensor Networks." Advanced Wireless Communication and Sensor Networks. Chapman and Hall/CRC 227-242.