IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 24, Issue 5, Ser. I (Sep. –Oct. 2022), PP 57-70
www.iosrjournals.org

Efficient Diagnosis of Atherosclerosis Through Combined Machine Learning And K-Means Clustering Methods

[1]Brajesh Kumar, Dr. [2]Md. Vaseem Naiyer, [3]Dr. B. Nageshwar Rao
[1]Research Scholar, Department of Computer Science and Engineering, Madhyanchal Professional University, Bhopal
[2]Supervisor, Department of Computer Science and Engineering, Madhyanchal Professional University, Bhopal
[3]Co Supervisor, Department of Computer Science and Engineering, Malla Reddy University, Hyderabad, Telangana, India.

Abstract *– The prediction of atherosclerosis disease is a very complex process. The complexity of prediction faces a problem of multiple attributes of heart disease-related symptoms. The machine learning algorithm plays an essential role in the prediction of atherosclerosis disease. The various authors and research scholar proposed various algorithms based on machine learning and artificial intelligence. This paper presents the study of machine learning algorithms for the prediction of atherosclerosis disease. The analysis of atherosclerosis disease applied four machine learning algorithms such as support vector machine, KNN, decision tree and naive Bayes. For the validation of algorithms applied four datasets: the applied dataset obtained from the UCI machine learning repository. For the evaluation of performance, measure three significant parameters like accuracy, specificity and sensitivity. I used MATLAB tools for all experimental research. MATLAB is algorithm analysis software. The analysis of performance suggests that a hybrid approach to integrate machine learning and k-means clustering is better than other machine learning algorithms. Coronary artery disease (CAD) is caused by atherosclerosis in coronary arteries and results in cardiac arrest and heart attack. For diagnosis of CAD, angiography is used which is a costly time consuming and highly technical invasive method. Researchers are, therefore, prompted for alternative methods such as machine learning algorithms that could use noninvasive clinical data for the disease diagnosis and assessing its severity. In this study, we present a novel hybrid method for CAD diagnosis; including risk factor identification using correlation-based feature subset (CFS) selection with particle swam optimization (PSO) search method and K- means clustering algorithms. Supervised learning algorithms such as multi- layer perceptron (MLP), multinomial logistic regression (MLR), fuzzy unordered rule induction algorithm (FURIA) and C4.5 are then used to model CAD cases.*

Keywords *– Atherosclerosis, Coronary Artery Disease, Machine Learning, K-Means Clustering, Particle Swarm Optimization, Feature Selection, Support Vector Machine, Clinical Decision Support System, Predictive Modeling, Artificial Intelligence in Healthcare.*

I.INTRODUCTION

The World Health Organization (WHO) reports that around 17.9 million people die each year from cardiovascular disease, and one-third of all premature deaths occur in people under the age of 70 [1]. The report also found that of the 17.3 million deaths from heart disease in 2020, approximately 6.2 million were from stroke and an estimated 7.3 million from coronary heart disease. The WHO predicted that around 23.6 million people would die from heart disease and stroke by 2030 [2]. Cardiovascular disease shows a number of symptoms including chest tightness, chest pressure, chest discomfort (angina), chest pain, shortness of breath, numbness, weakness or coldness in the legs or arms, if the blood vessels are constricted these parts of the body. Some other symptoms of this disease include nausea, fatigue, cold sweats, and discomfort in the neck, elbow, throat, jaw, left shoulder, upper abdomen, or back. However, the main causes of cardiovascular disease are age, smoking, sugar, obesity, depression, high blood pressure, high blood pressure, cholesterol, poor diet and lack of exercise [3]. Cardiovascular disease is also caused by damage to the coronary arteries, damage to all or part of the heart, or inadequate supply of nutrients and oxygen to the heart. There are various types of cardiovascular diseases, such as coronary artery disease, stroke, hypertensive heart disease, inflammatory heart disease, rheumatic heart disease, etc. [4]. Some types of cardiovascular disease, such as hypertrophic cardiomyopathy, dilated cardiomyopathy, and arrhythmogenic right ventricular cardiomyopathy, are genetically inherited. It is necessary to monitor a patient's cardiovascular symptoms to obtain medical advice from a healthcare professional. If cardiovascular disease is diagnosed at an early stage, the risk of death from the disease may be reduced. However, analyzing symptoms manually is difficult due to redundancy, multiple mapping, incompleteness, and a close relationship to

time in medical data. Additionally, finding the right medication for a patient after manually analyzing massive amounts of heart disease data is a significant challenge. To address this problem, machine learning (ML) technology helps create predictive models that can process and analyses large amounts of complex medical data and predict the absence or presence of cardiovascular disease for a patient with more accurate results. In the ML approach, a computer program is trained to perform a specific task in order to learn from its previous experience and predict the outcome of the test data based on the training data [5]. Machine learning techniques allow a machine to make correct decisions based on a built- in analytical model when it provides invisible data. Machine learning techniques take relatively less time to accurately predict. Therefore, an intelligent ML-based cardiovascular disease prediction system can help doctors make faster decisions so that they can offer medical treatments to a large number of patients in a short period of time; therefore, the ML model potentially saves millions of lives. Machine learning techniques have already achieved a significantly higher level of precision for classification-based problems [6]. Information abstraction has been achieved using several machines learning techniques, including feature selection, classification, and grouping [7], [8].

II.MOTIVATION

The main motivation of doing this research is to present a heart disease prediction model for the prediction of occurrence of heart disease. Further, this research work is aimed towards identifying the best classification algorithm for identifying the possibility of heart disease in a patient. This work is justified by performing a comparative study and analysis using three classification algorithms namely Naïve Bayes, Decision Tree, and Random Forest are used at different levels of evaluations. Although these are commonly used machine learning algorithms, the heart disease prediction is a vital task involving highest possible accuracy. Hence, the three algorithms are evaluated at numerous levels and types of evaluation strategies. This will provide researchers and medical practitioners to establish a better understanding and help them identify a solution to identify the best method for predicting the heart diseases.

III. OBJECTIVES OF THE STUDY

The report of WHO related to death in concern of heart disease is very scaring, the community of medical society and engineers put effort to predict early stage of heart disease for survival of human life. In the process of proposed methodology set objective given below

- Extensive analysis of different parameters of cardi vesicular disease using machine learning algorithms.
- Feature optimization for better processing of data using GSO and evaluation algorithms.
- Design ensemble-based classifier for the predication of CVD disease.
- Reduces training error during the sampling of data
- Reduces the feature separation process by principle of decomposition.
- To design and implement inter-features variance function for the processing of CVD data.

MACHINE LEARNING ALGORITHMS:

Data mining is a developing technique that is a direct result of the growing usage of computer databases to efficiently store and retrieve information. It involves the non-trivial extraction of innovative, implicit, and usable knowledge from big data sets. It also goes by the name Knowledge Discovery in Databases (KDD) and allows for the high- level, hypothesis-free data exploration, analysis, and visualization of enormous databases. By making predictions using a technique called modelling, it is possible to comprehend how data mining functions. Artificial neural networks, decision trees, and genetic algorithms are just a few of the data mining approaches that are the end result of years of study and product development. The technology behind data mining benefits from this ability to retrieve data as and when necessary. Data mining can be seen as a natural progression in the development of information technology. Huge volumes of data are readily available thanks to this technology, and the need to transform that data into knowledge and information is urgent. Data mining is the process of removing valuable information or patterns from vast amounts of data. Knowledge discovery (mining) in databases is sometimes referred to as KDD, knowledge extraction, data/pattern analysis, data dredging, information harvesting, business intelligence, and other names. The term "data mining" simply refers to the examination of data in a database using tools that look for trends or anomalies without knowledge of the data's meaning

IV. ATHEROSCLEROSIS:

One of the serious diseases that affects human arteries is atherosclerosis. The blood vessels that supply oxygen and blood to the heart and other body organs must originate from these arteries. These diseases are caused by too much fat, cholesterol, and other dangerous substances that can clog arteries and prevent the passage of blood and oxygen. Additionally, plagues are extremely harmful because they cause the arteries to become small and stiff. The most serious condition that can result in severe cardiac arrest is atherosclerosis. Retinal fundus imaging is useful for making an early diagnosis of this etiology. A sophisticated technique called fundus imaging

uses a 2-D representation of a 3-D retinal image that is slightly translucent. The FOV of fundus cameras may typically be extended up to 40 to 45 degrees, with an average pupil dilation of roughly 4 mm. These cameras designate a specific item for capturing photos of the eye's front region. The fundus is physically situated behind the eyeball. Thus, it is acceptable and possible to record the identification of normal alterations in the abnormal or diseased area. The use of a fundus camera can aid in early diagnosis, treatment process observation, and patient education.

Fundus cameras are used to take pictures that are utilized to diagnose and treat atherosclerosis, macular degeneration, and hypertension.

4.1 DATA MINING ALGORITHMS

Several data mining algorithms have been developed as a result of data mining research. These algorithms can be applied directly to a dataset to build models or to derive important conclusions and inferences from it. Decision tree, Naive Bayes, k- means, artificial neural network, etc. are some common data mining algorithms. The section that follows talks about them.

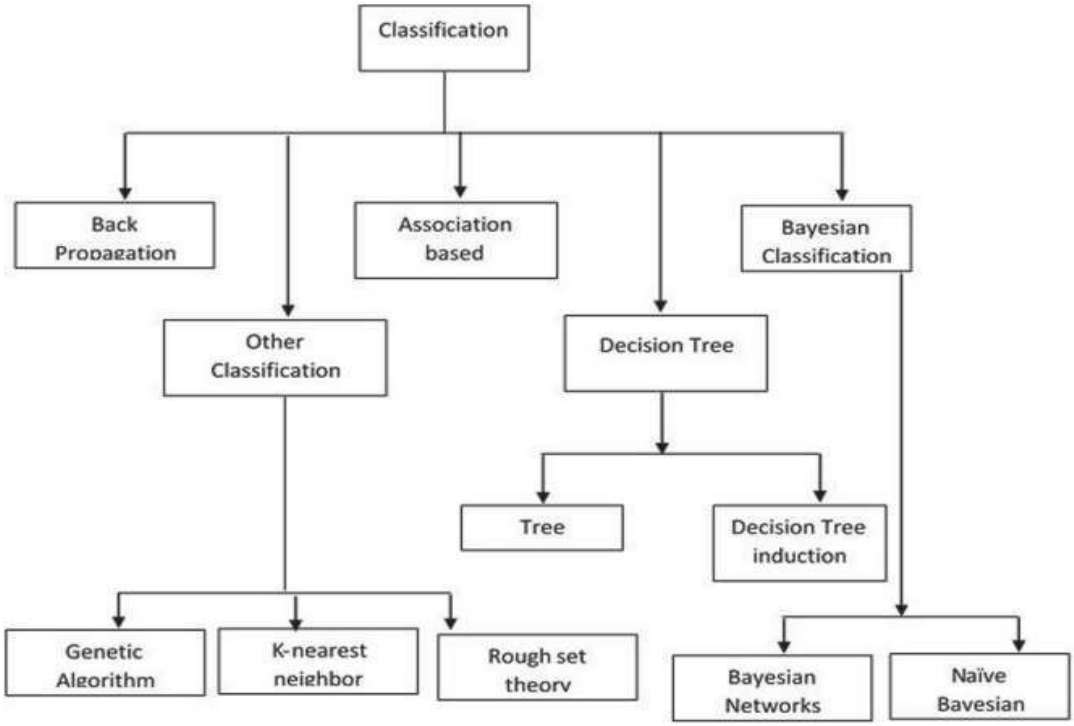


Fig 1: Data Mining Classification Method

4.1.1 DECISION TREE

A decision tree is a decision assistance tool that employs a graph or model resembling a tree of decisions and their potential outcomes, such as utility and chance event outcomes. It is one method of showing an algorithm. In operations research, decision trees are frequently used in decision analysis to assist in finding the approach most likely to succeed. It is also a well- liked machine learning tool. By sequentially mapping from the root node to the leaf nodes, a decision tree can be converted into a collection of rules with ease. Finally, by adhering to these guidelines, the proper conclusions can be drawn.

4.1.2 C4.5

A decision tree serves as the classifier in this instance. It is a supervised learning technique that employs pruning and knowledge acquisition to get better outcomes. It is widely used, relatively quick, and the output is simple to understand.

4.1.3 K-MEANS ALGORITHM

In order to make the members of a group more similar, K-means divides the provided set of objects into k groups. K- means does not only determine the number of clusters; it also "learns" the clusters on its own, without being told to which cluster a specific observation should belong. K-means can therefore be referred to as a semi-supervised learning technique. Over huge datasets, K-means is especially useful.

4.1.4 ID3 ALGORITHM

By comparing the values of the attributes, the ID3 algorithm (Quinlan 1986), a decision tree-building algorithm, determines the classification of objects. Starting with a collection of items and the specification of their properties, it creates the tree top down. A property is tested and the results used to partition the object at that node are set at each node of the tree. Until the set in a particular sub tree is homogeneous with regard to the classification criteria, this process is repeated recursively. It then turns into a leaf node. Information gain is maximized and entropy is reduced at each node. To put it another way, the quality that separates the candidate set into the subsets with the highest homogeneity is examined.

4.1.5 SUPPORT VECTOR MACHINE (SVM)

It is a method of supervised learning that divides data into two categories over a hyper plane. Except that it doesn't employ any Decision trees at all, Support vector machines carry out a similar duty to C4.5. To reduce the possibility of misclassification, the support vector machine tries to increase the margin (the distance between the hyper plane and the two nearest data points from each respective class). Scikit- learn, MATLAB, and LIBSVM are a few examples of well-known support vector machine implementations.

V. PROPOSED METHODOLOGY

5.1 PROPOSED ALGORITHM-1

The proposed ensemble classifier based on boosting methods. the support vector machine work as base classifier and KNN classifier work as variable classifier and collect the similar features of atherosclerosis data and process to support vector machine. The processing of features of atherosclerosis dataset by glow-worm optimization algorithm. The glow-worm optimization algorithm circulates the value of Lucifer in and collect relevant features to mapping of class. The process of ensemble classifier used K features and S classifier. The data search space of classifier is K ×S

Process of boosting ensemble classifier

The process of boosting

Consider K={ k1,k2,k3,… ,kn} be the atherosclerosis dataset

S={} ensemble classifier set C={c1,c2,c3,… ,cn} classifier set

$X = trainingset, X \in K$

$Y = test\ set, Y \in k$

M=n(K)

Consider init =1

L(init)= random subset of X; L(init)ϵX P(0)={}

Fori =1 to P do

if i>1

f(i) = irrelevant features of dataP(i-1) + M(i) P(i) = training process of classifier C(i) and S(i) S= S C(i)

for i = 1 toM R(i) = Y

next i

estimate = max(R(i): i=1,2, …, n)

Rule of ensemble classifier

1. the base class features = $\{k_1, k_2, \ldots \ldots kn\}$, where each $k_j = \{z_{j1}, z_{j2}, \ldots \ldots z_{js}\}$is collection of features of Em class $\{C_1, C_2, \ldots \ldots \ldots C_k\}$, the class of features are non-overlapping the process of ensemble describe as .

$Em(C_k) = \{\{z_j, z_k, \ldots \ldots . z_m\} \subseteq \{zf_1, zf_2, \ldots \ldots z_s\}\}$ (4.3)

2. let the kernel mapping of support vector machine classifier class is

$M_{i+1}[EM(C_g) = Em(C_g) \cup z_r] > S_i[Em(C_K)]$ (4.4)

3. prediction of ensemble classifier is $Em(C_k)$.

$Pre_{Em(C_K)} = pre(Em(C_K)) \leq S$ (4.5)

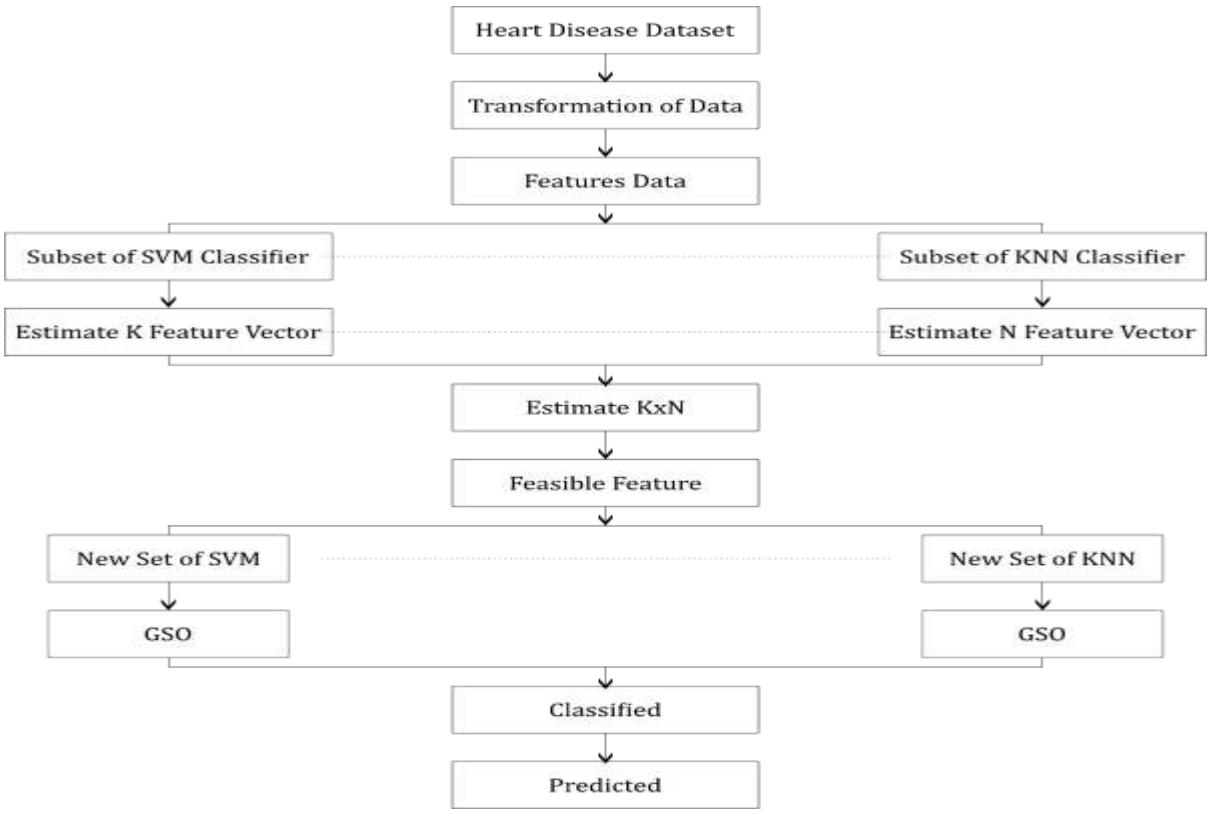


Figure .2: Process model of proposed work.

5.2 MACHINE LEARNING ALGORITHM

Machine learning also includes the examination of statistical models and methods. And computers employ these sorts of algorithms to fulfil a given purpose without the usage of obvious interruptions. It is a subfield in artificial intelligence. Machine learning algorithms use a mathematical model based on a training dataset to predict or make judgments without being explicitly programmed to do so. To complete the task effectively, machine learning employs several algorithms such as Random Forest, decision tree, and Navie Bayes. Machine learning is identical with statistical statistics, which uses computers to generate predictions. Data mining is the study of data in which machine learning focuses on examining and analyzing the data through unsupervised learning. Machine learning techniques, theory, and application fields are among the benefits that mathematical optimization provides. ML algorithms and approaches will process and predict data from massive amounts of data. First and foremost, the dataset, referred to as patient data collection, is pre-processed in this image. This step is required since the data has been cleansed in order for the machine learning algorithm to use it in this procedure. The model may use a single algorithm or a combination of algorithms in a hybrid approach.This model's output is a classifier; it is the most significant portion since it is the intelligence, and classifiers are the part that permits prediction. If the classifier is given the input data, it can predict without the need for human intervention. If the input data is information about a new patient and the data-set fed into the model is a medical dataset of healthy and unwell patients.

5.3 SUPPORT VECTOR MACHINE (SVM)

SVM (Support vector machine) is machine learning algorithm derived by Vipin in 1990[10]. The support vector machine applied in various filed of image classification and pattern recognition. The nature of support vector machine is linear, non- linear and sigmoid. The non- linear support vector machine mapping the feature data with respect to one plane to another plan [2, 3, 16]. The separation of data plan is non- linear and decision factor correlate with margin function of support vector. The hyperplane of equation is derived as

$WD.xi + b \geq 1\ if\ yi = 1$ (4.6)

$WD.xi + b \leq -1\ if\ yi = -1$

Here W is weight vector, x is input vector yi label o class and b is bias.

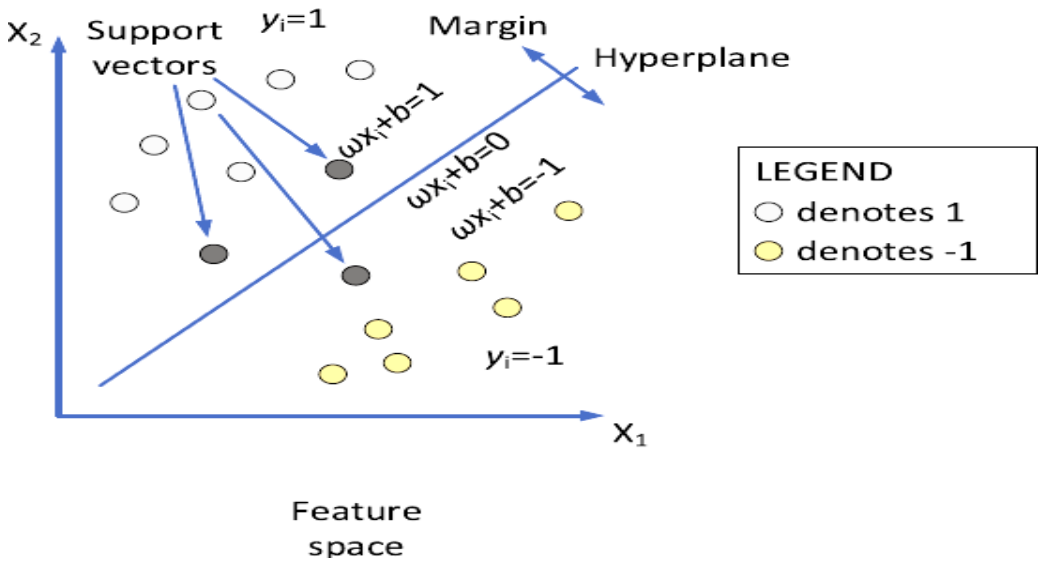


Figure 3: Process block diagram of support vector machine.

5.4 KNN

The KNN classifier is simple algorithm of machine learning, it also knows as lazy classifier. The classification accuracy of KNN classifier varies in range of 70-80%. The major utility of KNN classifier in case of pattern recognition. The KNN classification algorithm applied on the case of continuous nature of attribute. The processing of KNN algorithm describe here

1. Estimate K training attribute which belong to unknown attribute
2. Chose the common occurring classification of K

For the estimation of similarity in class of K instance applied different distance equation. The very famous distance equation is Euclidean distance equation.

Input: A data set according to sample selection Output : a mixed transform table data
class: E={},the set of the equivalence classes QIC={},set of equivalence classes with similar QI sets
CIP{}, set of attributes with similar class
DIP=number of different class values in the remaining dataset Begin
While CIP>= attribute
Cluster T to m tables according QI For i=1 to m
Bucketize attributes according SA values While |DIPi|>=ℓ Create_equivalence_classes ()
E=E U Create_equivalence_classes() return E
Incorporate the remaining attributes to E End
Generate equivalence class with prototype is Input: CIP
Output :E Begin
Randomly selection of a attributes tm from the smallest group
E={tm}
For p=1 until attribute-1
Select a attributes tp that minimizes the gcp E=EU tp
Remove tp from T Remove tm from T Return E
End
Process of cluster generation in prototype classification Input: data set used defined
Output: QIC={}, set of tables with attributes with similar QI sets Begin
Insert T to the decision tree classification QIC={ QIC1, QIC2,… QICm }
return QIC End

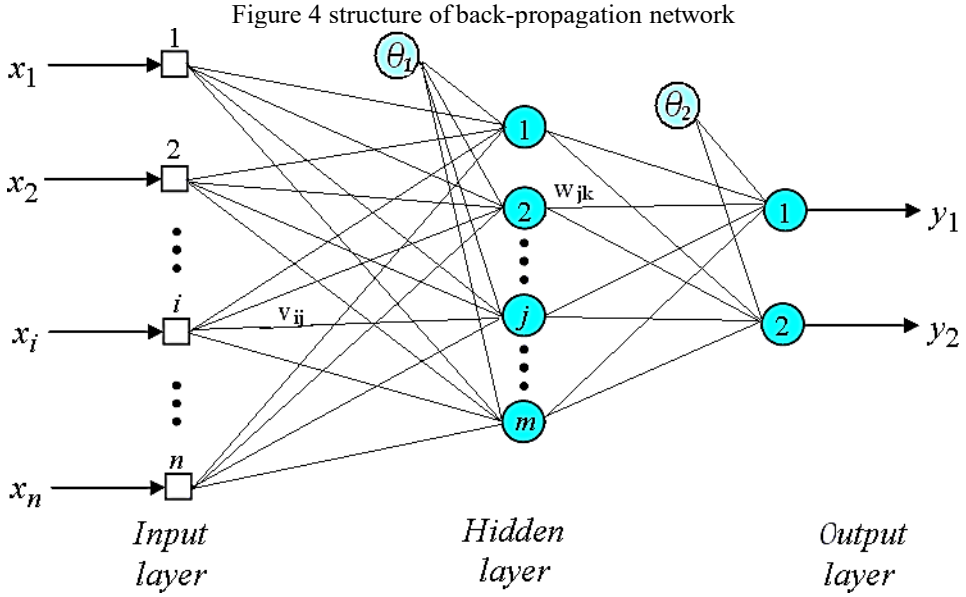
5.5 DECISION TREE (DT) ALGORITHM

A decision tree is predictive model of data classification. The p rocessing of decision tree algorithm follows the tree structure of feature set. The partition of branch applied the concept of maximum entropy. The feature attribute has maximum entropy treat as leaf node class. The learning process of decision tree algorithm is very simple and suitable for the prediction of medical data analysis. The decision tee rule formation according to the behaviors of feature attribute narrated in terms of maximum occurrence of probability. The core algorithm of decision tree is C4.5. the processing of these algorithm on design of tree in manner of top to down as class label. The entropy and gain are two functions manage the processing of decision tree algorithm.

5.6 PROPOSED ALGORITHM-2

The back propagation neural network model basically controls the rate of error in terms of desired output difference proceed in source of network and enhance the rate of detection of given sample data of breast cancer. The processing part of back propagation algorithms divided into two part one is training part and other is application part.

Figure 4 structure of back-propagation network



5.7 PROPOSED ALGORITHM-3

Proposed models are creating for data training for minority and majority class data sample for processing of associative classification level of rules. The associative classification process a data input for training phase for SMOTE and CMTNN sampling technique for classifier. While single- layer SOM networks can potentially learn virtually any input output relationship, SOM networks with single layers might learn complex relationships more quickly [12]. The function SOM creates wiener and successor matrix. For example, a ensemble layer network has connections from layer 1 to layer 2, layer 2 to layer 3, and layer 1 to layer 3. The ensemble - layer network also has connections from the input to all cascaded layers. The additional connections
might improve the speed at which the network learns the desired relationship. SOM artificial intelligence model is similar to feed- forward back-propagation neural network in using the back-propagation algorithm for weights updating, but the main symptom of this network is that each layer of neurons related to all previous layer of neurons. Tan-sigmoid transfer function, log - sigmoid transfer function and pure linear threshold functions were used to reach the optimized status.

5.5.1.1 Data are passes through ACR

5.5.1.2 ACR makes a multi-level rule set using rule mining algorithm

5.5.1.3 level of rules going to SOM ensemble process

5.5.1.4 The training phase data are passes through SMOTE AND CMTNN sampler

5.5.1.5 The sampling of data passes through SOM AND balanced the data for minority and majority ratio of class

5.5.1.6 The sampled data assigned to k-type binary class

5.5.1.7 Binary class data are coded in bit form

5.5.1.8 if code bit value is single assigned the class value

5.5.1.9 Else data goes to training phase

5.5.1.10 . Balanced part of training is updated

5.5.1.11 Find accuracy and relative mean Error

5.5.1.12 Exit

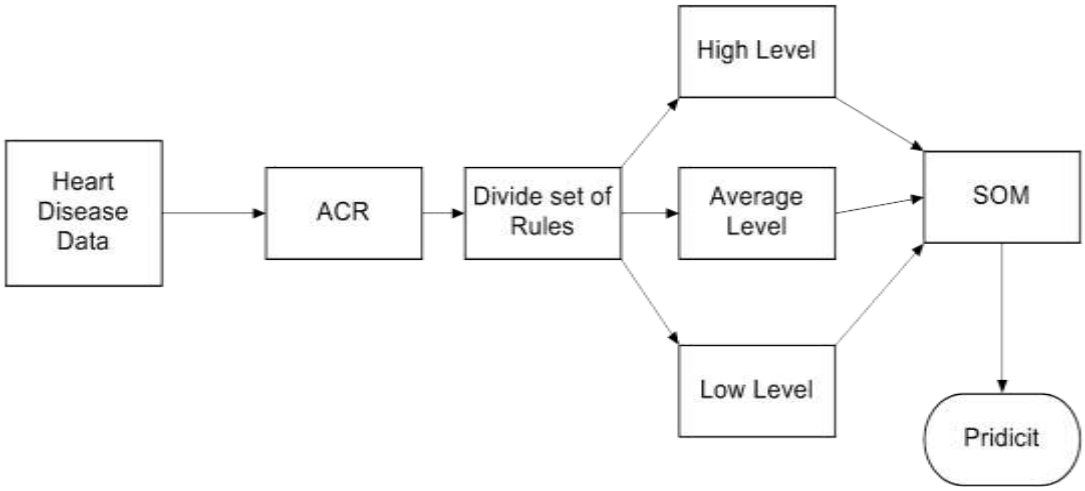


Figure 5 proposed model for ensemble based associative classification with SOM network

VI. IMPLEMENTATATION PROCCESS

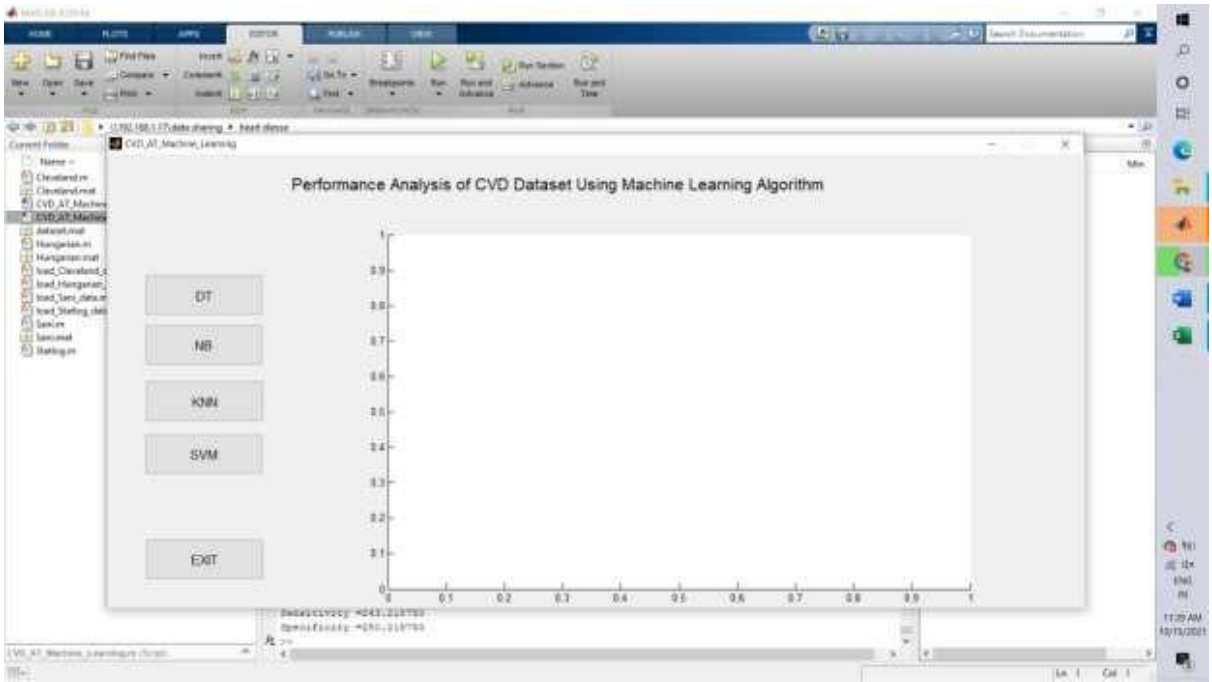


Figure 6 : initial window of the simulation in MATLAB software and here we can seen the GUI of our model .

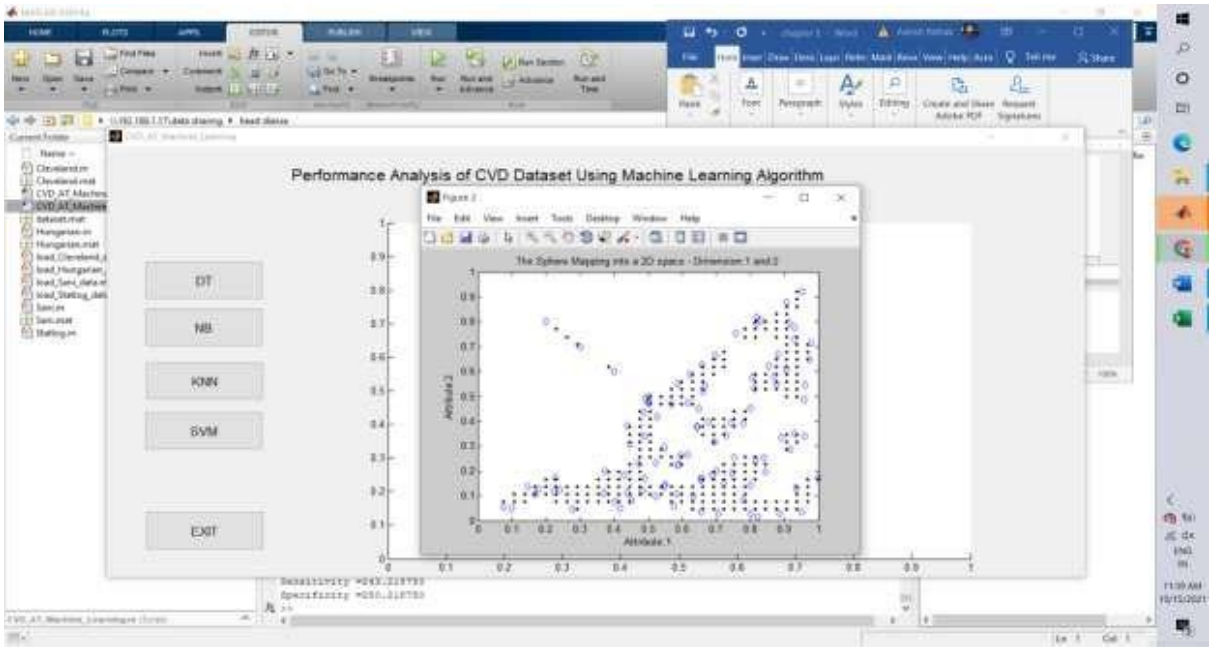


Figure 7: window show that using the DT method simulation in MATLAB software.

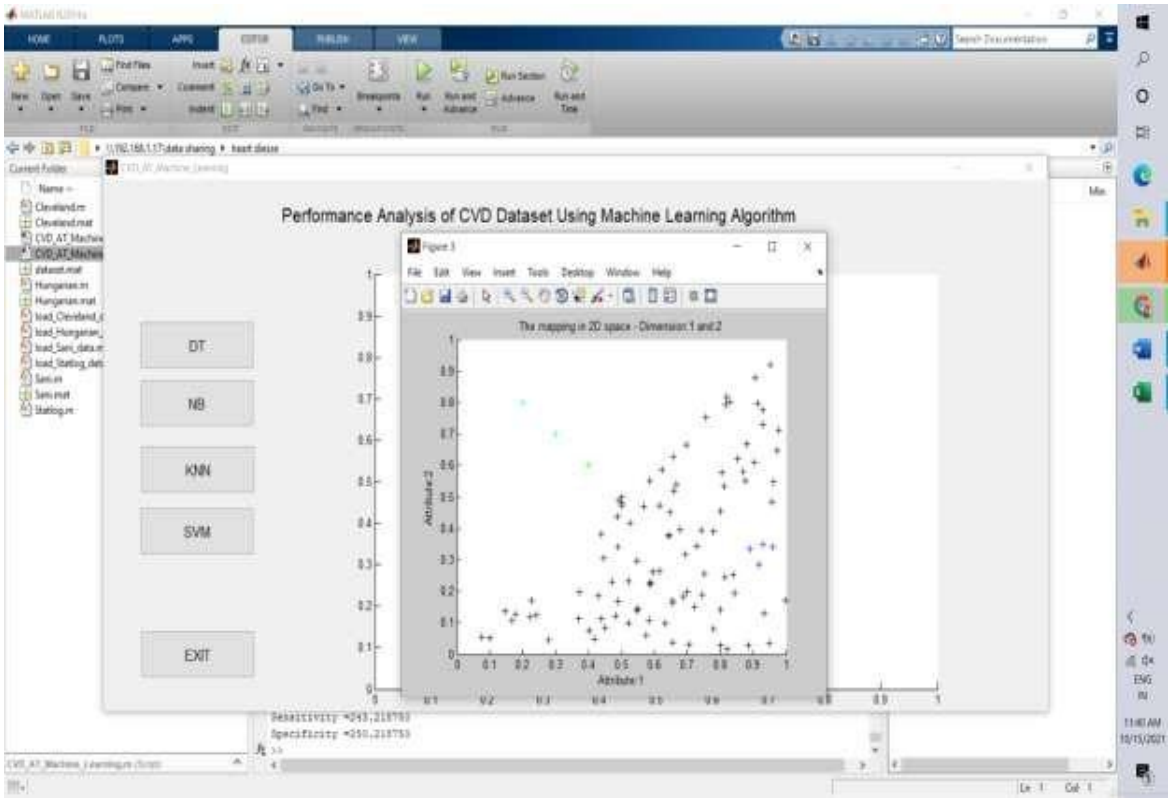


Figure 8.: window show that the of the out figure 3 using the DT method simulation in MATLAB software.

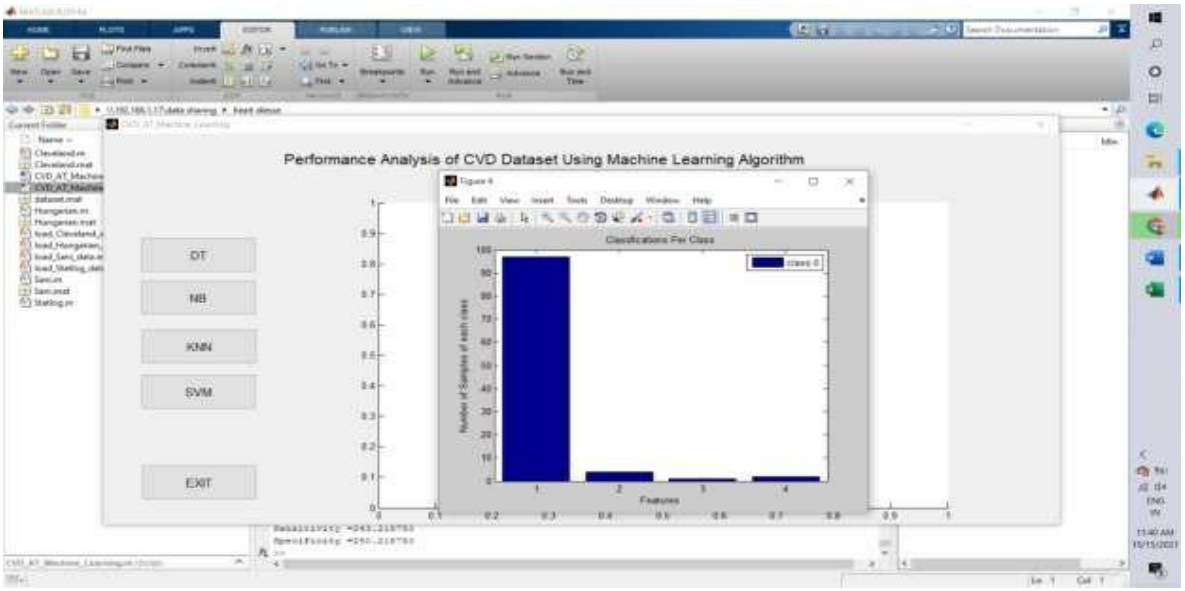


Figure 9 : window show that the of the out figure 4 using the DT method simulation in MATLAB software.

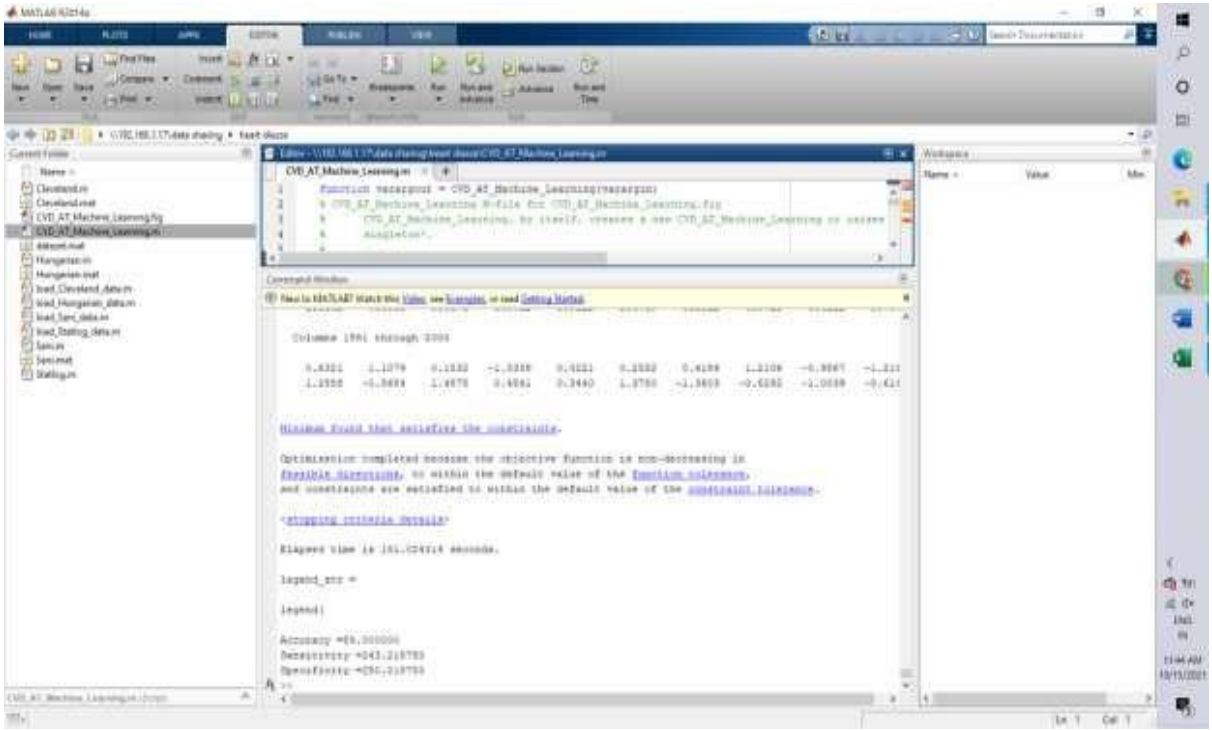


Figure 10 : window show that the of the output result of accuracy, sensitivity, specificity using DT method simulation in MATLAB software.

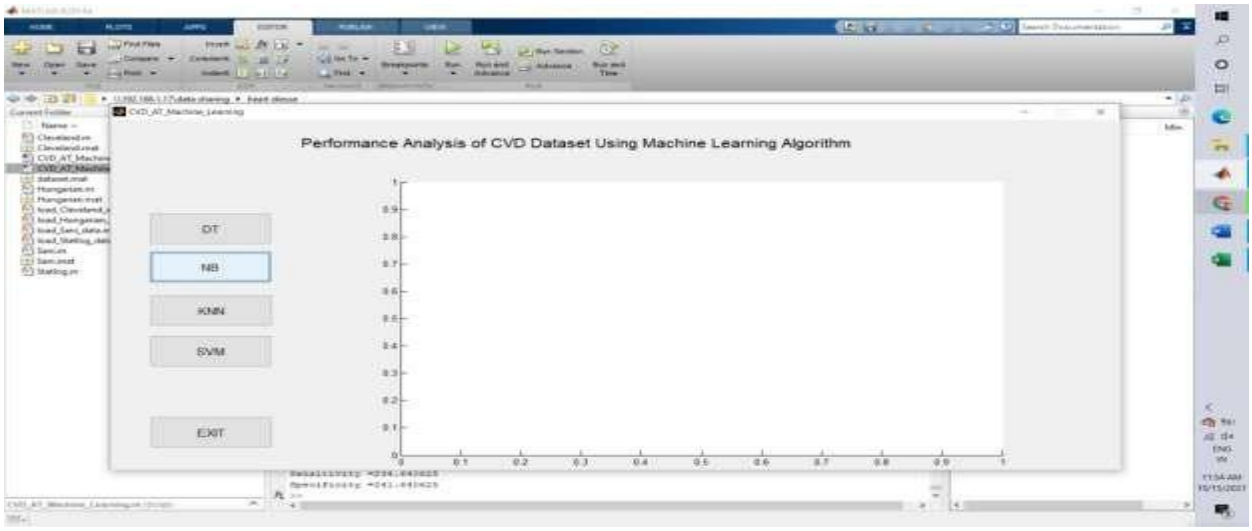


Figure 1 1 : window show that the GUI of our simulation model hit the NB method n MATLAB software.

VII. CONCLUSION

The early detection of cardiovascular disease (CVD) saves millions of lives worldwide. Cardiovascular disease has various causes high blood pressure, diabetes, high cholesterol and many other factors. Atherosclerosis disease is the leading cause of artery arrest and heart stroke. The performance of automatic CVD detection is a very challenging task. The accurate detection of disease symptoms suggests medicine and curing heart disease. Early data mining algorithms are applied to detect and classify atherosclerosis disease. The sampling and selection of attributes face a bottleneck problem in classifying heart disease. The advancement of machine learning algorithms improves the classification and detection ratio of heart disease. The prediction and analysis of atherosclerosis disease machine learning applied four classification algorithms: vector machine, decision tree, naïve Bayes and KNN. The results of the support vector machine influence the process of atherosclerosis disease detection. Four data sets validate the algorithm used: Cleveland, Hungarian, Starlog and Z-Alizadeh Sani. The training and testing process applied a 10 cross- fold ratio. The NB and DT algorithms give better performance than KNN algorithms.The accuracy of KNN algorithms depends on the variation of K attributes. The selection of features and attributes plays a significant role in prediction accuracy. The machine learning algorithm's outcomes reached an overall accuracy of 88% and 85% of sensitivity. Decision tree, NB, and KNN were the accuracies compared with the support vector machine, respectively 87.00%, 89.01%, and 81.10%. Thus, this comparison has shown that the support vector machine is better. Still accuracy of machine learning algorithms is a challenging task. A large number of features and the dataset's size decline the classification algorithms' performance. The glowworm optimization feature selection cum optimization algorithm reduces the number of features and increases the atherosclerosis disease detection classification ratio. The glowworm optimization reduces the unwanted features of datasets. The proposed ensemble classifier overcomes the limitation of the binary classification algorithm. The ensemble classifier used two classifier support vector machines and KNN. The process of boosting increases the mapping of features and reduces the noise and outlier data of atherosclerosis patients. The proposed algorithm was assessed on different datasets and tried with a combination of datasets.

VIII. SUGGESTION FOR FUTURE WORK

In the current decade, the cardiovascular disease classification application is not commercial. However, prosed algorithms have applied any classification for different heart disease. Therefore, the specific limitation of the proposed algorithms handles in the future is mentioned here.

1. Development of a sensory algorithm for quality data sampling.
2. Minimization of null value and normalization of data of user.
3. lightweight feature extraction methods
4. Generalization of results in all tasks of CVD classification.

REFERENCES

[1]. Terrada, Oumaima, BouchaibCherradi, Abdelhadi Raihani, and Omar Bouattane. "A novel medical diagnosis support system for predicting patients with atherosclerosis diseases." *Informatics in Medicine Unlocked* 21 (2020): 100483.
[2]. Terrada, Oumaima, BouchaibCherradi, Abdelhadi Raihani, and Omar Bouattane. "Classification and Prediction of atherosclerosis diseases using machine learning algorithms." In *2019 5th International Conference on Optimization and Applications (ICOA)*, pp. 1-5. IEEE, 2019.
[3]. Yilmaz, Nihat, OnurInan, and Mustafa SerterUzer. "A new data preparation method based on clustering algorithms for diagnosis systems

of heart and diabetes diseases." *J Med Syst* 38 (2014): 48-59.
[4]. Bhatla, Nidhi, and Kiran Jyoti. "A Novel Approach for heart disease diagnosis using Data Mining and Fuzzy logic." *International Journal of Computer Applications* 54, no. 17 (2012).
[5]. Munger, Eric, John W. Hickey, Amit K. Dey, Mohsin Saleet Jafri, Jason M. Kinser, and Nehal N. Mehta. "Application of machine learning in understanding atherosclerosis: Emerging insights." *APL bioengineering* 5, no. 1 (2021): 011505.
[6]. Parameswari, C., and S. Siva Ranjani. "Prediction of atherosclerosis pathology in retinal fundal images with machine learning approaches." *Journal of Ambient Intelligence and Humanized Computing* (2020): 1-11.
[7]. Kolossváry, Márton, Carlo N. De Cecco, Gudrun Feuchtner, and PálMaurovich-Horvat. "Advanced atherosclerosis imaging by CT: radiomics, machine learning and deep learning." *Journal of cardiovascular computed tomography* 13, no. 5 (2019): 274-280.
[8]. Razzak, Muhammad Imran, Muhammad Imran, and Guandong Xu. "Big data analytics for preventive medicine." *Neural Computing and Applications* 32, no. 9 (2020): 4417-4451.
[9]. Singh, Navdeep, Punjab Firozpur, and Sonika Jindal. "Heart disease prediction system using hybrid technique of data mining algorithms." *International Journal of Advance Research, Ideas and Innovations in Technology* 4, no. 2 (2018): 982-987.
[10]. Rao, V. Sree Hari, and M. Naresh Kumar. "Novel approaches for predicting risk factors of atherosclerosis." *IEEE journal of biomedical and health informatics* 17, no. 1 (2012): 183-189.
[11]. Qawqzeh, Yousef K., Mohammad Mahmood Otoom, Fayez Al-Fayez, Ibrahim Almarashdeh, MutasemAlsmadi, and GhaithJaradat. "A Proposed Decision Tree Classifier for Atherosclerosis Prediction and Classification." *IJCSNS* 19, no. 12 (2019): 197.
[12]. AnimeshHazra, Subrata Kumar Mandal, Amit Gupta, Arkomita Mukherjee and Asmita Mukherjee. "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review." *Advances in Computational Sciences and Technology* 10, no. 7 (2017): 2137-2159.
[13]. Shouman, Mai, Tim Turner, and Rob Stocker. "Integrating Naive Bayes and K-means clustering with different initial centroid selection methods in the diagnosis of heart disease patients." *CS & IT-CSCP* (2012): 125-137.
[14]. Garg, Ravi, Elissa Oh, Andrew Naidech, Konrad Kording, and Shyam Prabhakaran. "Automating ischemic stroke subtype classification using machine learning and natural language processing." *Journal of Stroke and Cerebrovascular Diseases* 28, no. 7 (2019): 2045-2051.
[15]. Terrada, Oumaima, BouchaibCherradi, Abdelhadi Raihani, and Omar Bouattane. "Classification and Prediction of atherosclerosis diseases using machine learning algorithms." In *2019 5th International Conference on Optimization and Applications (ICOA)*, pp. 1-5. IEEE, 2019.
[16]. Han, Donghee, Kranthi K. Kolli, Subhi J. Al'Aref, Lohendran Baskaran, Alexander R. van Rosendael, Heidi Gransar, Daniele Andreini et al. "Machine learning framework to identify individuals at risk of rapid progression of coronary atherosclerosis: from the PARADIGM registry." *Journal of the American Heart Association* 9, no. 5 (2020): e013958.
[17]. Nikan, Soodeh, FemidaGwadry-Sridhar, and Michael Bauer. "Machine learning application to predict the risk of coronary artery atherosclerosis." In *2016 International conference on computational science and computational intelligence (CSCI)*, pp. 34-39. IEEE, 2016.
[18]. Serrano, José Ignacio, M. Tomeckova, and Jana Zvárová. "Machine learning methods for knowledge discovery in medical data on atherosclerosis." *European Journal for Biomedical Informatics* 2, no. 1 (2006): 6-33.
[19]. Shanmugasundaram, G., V. Malar Selvam, R. Saravanan, and S. Balaji. "An Investigation of Heart Disease Prediction Techniques." In *2018 IEEE International Conference on System, Computation, Automation and Networking (ICSCA)*, pp. 1-6. IEEE, 2018.
[20]. Georga, Eleni I., Nikolaos S. Tachos, Antonis I. Sakellarios, Vassiliki I. Kigka, Themis P. Exarchos, Gualtiero Pelosi, OberdanParodi, Lampros K. Michalis, and Dimitrios I. Fotiadis. "Artificial intelligence and data mining methods for cardiovascular risk prediction." In *Cardiovascular Computing—Methodologies and Clinical Applications*, pp. 279-301. Springer, Singapore, 2019.
[21]. Kumar, Abhishek, Pardeep Kumar, Ashutosh Srivastava, VD Ambeth Kumar, K. Vengatesan, and Achintya Singhal. "Comparative Analysis of Data Mining Techniques to Predict Heart Disease for Diabetic Patients." In *International Conference on Advances in Computing and Data Sciences*, pp. 507-518. Springer, Singapore, 2020.
[22]. Magesh, G., and P. Swarnalatha. "Optimal feature selection through a cluster-based DT learning (CDTL) in heart disease prediction." *Evolutionary Intelligence* (2020): 1-11.
[23]. Rémy, NfongourainMougnutou, TekinzangTedondjio Martial, and TayouDjamegniClémentin. "The prediction of good physicians for prospective diagnosis using data mining." *Informatics in medicine unlocked* 12 (2018): 120-127.
[24]. Krishnan, Santhana, and S. Geetha. "Prediction of Heart Disease Using Machine Learning Algorithms." In *2019 1st international conference on innovations in information and communication technology (ICIICT)*, pp. 1-5. IEEE, 2019.
[25]. Repaka, Anjan Nikhil, Sai Deepak Ravikanti, and Ramya G. Franklin. "Design and implementing heart disease prediction using naives Bayesian." In *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 292-297. IEEE, 2019.
[26]. Repaka, Anjan Nikhil, Sai Deepak Ravikanti, and Ramya G. Franklin. "Design and implementing heart disease prediction using naives Bayesian." In *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 292-297. IEEE, 2019.
[27]. Kazerouni, Faranak, AzadehBayani, FarkhondehAsadi, Leyla Saeidi, Nasrin Parvizi, and Zahra Mansoori. "Type2 diabetes mellitus prediction using data mining algorithms based on the long-noncoding RNAs expression: a comparison of four data mining approaches." *BMC bioinformatics* 21, no. 1 (2020): 1-13.
[28]. Shaji, Shaicy P. "Predictionand Diagnosis of Heart Disease Patients using Data Mining Technique." In *2019 international conference on communication and signal processing (ICCSP)*, pp. 0848-0852. IEEE, 2019.
[29]. Mathan, K., PriyanMalarvizhi Kumar, Parthasarathy Panchatcharam, Gunasekaran Manogaran, and R. Varadharajan. "A novel Gini index decision tree data mining method with neural network classifiers for prediction of heart disease." *Design automation for embedded systems* 22, no. 3 (2018): 225-242.
[30]. Atallah, Rahma, and Amjed Al-Mousa. "Heart Disease Detection Using Machine Learning Majority Voting Ensemble Method." In *2019 2nd International Conference on new Trends in Computing Sciences (ICTCS)*, pp. 1-6. IEEE, 2019.
[31]. Ayatollahi, Haleh, Leila Gholamhosseini, and Masoud Salehi. "Predicting coronary artery disease: a comparison between two data mining algorithms." *BMC public health* 19, no. 1 (2019): 1-9.
[32]. Bashir, Saba, Zain Sikander Khan, Farhan Hassan Khan, Aitzaz Anjum, and Khurram Bashir. "Improving heart disease prediction using feature selection approaches." In *2019 16th international bhurban conference on applied sciences and technology (IBCAST)*, pp. 619-623. IEEE, 2019.
[33]. Bashir, Saba, Zain Sikander Khan, Farhan Hassan Khan, Aitzaz Anjum, and Khurram

Bashir. "Improving heart disease prediction using feature selection approaches." In *2019 16th international bhurban conference on applied sciences and technology (IBCAST)*, pp. 619-623. IEEE, 2019.
[34]. Amin, Mohammad Shafenoor, Yin Kia Chiam, and Kasturi DewiVarathan. "Identification of significant features and data mining techniques in predicting heart disease." *Telematics and Informatics* 36 (2019): 82-93.
[35]. Dahiwade, Dhiraj, Gajanan Patle, and EktaaMeshram. "Designing disease prediction model using machine learning approach." In *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 1211-1215. IEEE, 2019.
[36]. Deepika, M., and K. Kalaiselvi. "A Empirical study on disease diagnosis using data mining techniques." In *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, pp. 615-620. IEEE, 2018.
[37]. Dhar, Sanchayita, Krishna Roy, Tanusree Dey, Pritha Datta, and Ankur Biswas. "A hybrid machine learning approach for prediction of heart diseases." In *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, pp. 1-6. IEEE, 2018.
[38]. Dinesh, Kumar G., K. Arumugaraj, Kumar D. Santhosh, and V. Mareeswari. "Prediction of cardiovascular disease using machine learning algorithms." In *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, pp. 1-7. IEEE, 2018.
[39]. Dwivedi, Ashok Kumar. "Performance evaluation of different machine learning techniques for prediction of heart disease." *Neural Computing and Applications* 29, no. 10 (2018): 685-693.
[40]. Rathnayakc, BandarageShehaniSanketha, and Gamage UpekshaGanegoda. "Heart diseases prediction with data mining and neural network techniques." In *2018 3rd International Conference for Convergence in Technology (I2CT)*, pp. 1-6. IEEE, 2018.
[41]. Goel, Sakshi, Abhinav Deep, Shilpa Srivastava, and Aprna Tripathi. "Comparative Analysis of various Techniques for Heart Disease Prediction." In *2019 4th International Conference on Information Systems and Computer Networks (ISCON)*, pp. 88-94. IEEE, 2019.
[42]. Haq, Amin Ul, Jian Ping Li, Muhammad Hammad Memon, Shah Nazir, and Ruinan Sun. "A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms." *Mobile Information Systems* 2018 (2018).
[43]. Hasan, KM Zubair, Shourob Datta, Md Zahid Hasan, and Nusrat Zahan. "Automated prediction of heart disease patients using sparse discriminant analysis." In *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, pp. 1-6. IEEE, 2019.
[44]. Hossain, Rifat, SM Hasan Mahmud, Md AltabHossin, Sheak Rashed Haider Noori, and Hosney Jahan. "PRMT: Predicting Risk Factor of Obesity among Middle-Aged People Using Data Mining Techniques." *Procedia computer science* 132 (2018): 1068-1076.
[45]. Maji, Srabanti, and Srishti Arora. "Decision tree algorithms for prediction of heart disease." In *Information and communication technology for competitive strategies*, pp. 447-454. Springer, Singapore, 2019.
[46]. Latha, C. Beulah Christalin, and S. Carolin Jeeva. "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques." *Informatics in Medicine Unlocked* 16 (2019): 100203.
[47]. Maragatham, G., and Shobana Devi. "LSTM model for prediction of heart failure in big data." *Journal of medical systems* 43, no. 5 (2019): 1-13.
[48]. Narayan, Subhashini, and E. Sathiyamoorthy. "A novel recommender system based on FFT with machine learning for predicting and identifying heart diseases." *Neural Computing and Applications* 31, no. 1 (2019): 93-102.
[49]. Rady, El-Houssainy A., and Ayman S. Anwar. "Prediction of kidney disease stages using data mining algorithms." *Informatics in Medicine Unlocked* 15 (2019): 100178.
[50]. Raju, Cincy, E. Philipsy, Siji Chacko, L. Padma Suresh, and S. Deepa Rajan. "A survey on predicting heart disease using data mining techniques." In *2018 conference on emerging devices and smart systems (ICEDSS)*, pp. 253-255. IEEE, 2018.
[51]. Ramasamy, S., and K. Nirmala. "Disease prediction in data mining using association rule mining and keyword based clustering algorithms." *International Journal of Computers and Applications* 42, no. 1 (2020): 1-8.
[52]. Raza, Khalid. "Improving the prediction accuracy of heart disease with ensemble learning and majority voting rule." In *U-Healthcare Monitoring Systems*, pp. 179-196. Academic Press, 2019.
[53]. Safdar, Saima, Saad Zafar, Nadeem Zafar, and Naurin Farooq Khan. "Machine learning based decision support systems (DSS) for heart disease diagnosis: a review." *Artificial Intelligence Review* 50, no. 4 (2018): 597-623.
[54]. Sarkar, Bikash Kanti. "Hybrid model for prediction of heart disease." *Soft Computing* 24, no. 3 (2020): 1903-1925.
[55]. Singh, Poornima, Sanjay Singh, and Gayatri S. Pandi-Jain. "Effective heart disease prediction system using data mining techniques." *International journal of nanomedicine* 13, no. T-NANO 2014 Abstracts (2018): 121.
[56]. Woldemichael, FikirteGirma, and Sumitra Menaria. "Prediction of diabetes using data mining techniques." In *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 414-418. IEEE, 2018.
[57]. Wu, Han, Shengqi Yang, Zhangqin Huang, Jian He, and Xiaoyi Wang. "Type 2 diabetes mellitus prediction model based on data mining." *Informatics in Medicine Unlocked* 10 (2018): 100-107.
[58]. Palaniappan, Sellappan, and Rafiah Awang. "Intelligent heart disease prediction system using data mining techniques." In *2008 IEEE/ACS international conference on computer systems and applications*, pp. 108-115. IEEE, 2008.
[59]. Yekkala, Indu, and Sunanda Dixit. "Prediction of heart disease using rando m forest and rough set based feature selection." *International Journal of Big Data and Analytics in Healthcare (IJBDAH)* 3, no. 1 (2018): 1-12.
[60]. Hasan, S. M. M., M. A. Mamun, M. P. Uddin, and M. A. Hossain. "Comparative analysis of classification approaches for heart disease prediction." In *2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)*, pp. 1-4. IEEE, 2018.
[61]. Kolukısa, Burak, HilalHacılar, Mustafa Kuş, BurcuBakır-Güngör, Atilla Aral, and VehbiÇağrıGüngör. "Diagnosis of Coronary Heart Disease via Classification Algorithms and a New Feature Selection Methodology." *International Journal of Data Mining Science* 1, no. 1 (2019): 8-15.
[62]. Anitha, S., and N. Sridevi. "Heart disease prediction using data mining techniques." *Journal of Analysis and Computation* (2019).
[63]. Diwakar, Manoj, Amrendra Tripathi, Kapil Joshi, Minakshi Memoria, and Prabhishek Singh. "Latest trends on heart disease prediction using machine learning and image fusion." *Materials Today: Proceedings* 37 (2021): 3213-3218.
[64]. Malav, Amita, and Kalyani Kadam. "A hybrid approach for heart disease prediction using artificial neural network and K-means." *International Journal of Pure and Applied Mathematics* 118, no. 8 (2018): 103-10.
[65]. Ismail, Ahmed, Samir Abdlerazek, and I. M. El- Henawy. "Big data analytics in heart diseases prediction." *Journal of Theoretical and Applied Information Technology* 98, no. 11 (2020).
[66]. Kefelegn, Shambel, and Pooja Kamat. "Prediction and analysis of liver disorder diseases by using data mining technique: survey." *International Journal of Pure and Applied Mathematics* 118, no. 9 (2018): 765-770.
[67]. Domadiya, N ikunj, and Udai Pratap Rao. "Privacy-preserving association rule mining for horizontally partitioned healthcare data: a case study on the heart diseases." *Sādhanā* 43, no. 8 (2018): 1-9.
[68]. Reddy, N. Satish Chandra, Song Shue Nee, Lim Zhi Min, and Chew Xin Ying. "Classification and feature selection approaches by

machine learning techniques: Heart disease prediction." *International Journal of Innovative Computing* 9, no. 1 (2019).
[69]. Awan, Shahid Mehmood, Muhammad Usama Riaz, and Abdul Ghaffar Khan. "Prediction of heart disease using artificial neural network." *VFAST Transactions on Software Engineering* 6, no. 1 (2018): 51-61.
[70]. Sumathi, A., and T. T. Abiraami. "Prediction of heart disease complication for diabetic patient using data mining techniques." *Int J Pure Appl Math* 12 (2018): 13869-13879.
[71]. Kodati, Sarangam, and R. Vivekanandam. "Analysis of heart disease using in data mining tools Orange and Weka." *Global Journal of Computer Science and Technology* 18, no. 1 (2018).
[72]. Tuli, Shreshth, N ipamBasumatary, Sukhpal Singh Gill, Mohsen Kahani, Rajesh Chand Arya, Gurpreet Singh Wander, and Rajkumar Buyya. "HealthFog: An ensemble deep learning based Smart Healthcare System for Automatic Diagnosis of Heart Diseases in integrated IoT and fog computing environments." *Future Generation Computer Systems* 104 (2020): 187-200.
[73]. Desai, Shrinivas D., ShantalaGiraddi, Prashant Narayankar, Neha R. Pudakalakatti, and Shreya Sulegaon. "Back-propagation neural network versus logistic regression in heart disease classification." In *Advanced computing and communication technologies*, pp. 133-144. Springer, Singapore, 2019.
[74]. Pavithra, M., A. M. Sindhana, T. Subajanaki, and S. Mahalakshmi. "Effective Heart Disease Prediction Systems Using Data Mining Techniques." *Annals of the Romanian Society for Cell Biology* (2021): 6566-6571.
[75]. Khennou, Fadoua, Charif Fahim, Habiba Chaoui, and Nour El HoudaChaoui. "A machine learning approach: Using predictive analytics to identify and analyze high risks patients with heart disease." *International Journal of Machine Learning and Computing* 9, no. 6 (2019): 762-767.
[76]. Ali, Liaqat, AwaisN iamat, Javed Ali Khan, Noorbakhsh Amiri Golilarz, XiongXingzhong, Adeeb Noor, Redhwan Nour, and Syed Ahmad Chan Bukhari. "An optimized stacked support vector machines based expert system for the effective prediction of heart failure." *IEEE Access* 7 (2019): 54007-54014.
[77]. Mohan, Senthilkumar, ChandrasegarThirumalai, and Gautam Srivastava. "Effective heart disease prediction using hybrid machine learning techniques." *IEEE access* 7 (2019): 81542-81554.
[78]. Qawqzeh, Yousef K., Mohammad Mahmood Otoom, Fayez Al-Fayez, Ibrahim Almarashdeh, MutasemAlsmadi, and GhaithJaradat. "A Proposed Decision Tree Classifier for Atherosclerosis Prediction and Classification." *IJCSNS* 19, no. 12 (2019): 197.
[79]. Chaurasia, Vikas, Saurabh Pal, and B. B. Tiwari. "Prediction of benign and malignant breast cancer using data mining techniques." *Journal of Algorithms & Computational Technology* 12, no. 2 (2018): 119-126.
[80]. Samuel, Oluwarotimi Williams, Bin Yang, YanjuanGeng, Mojisola Grace Asogbon, Sandeep Pirbhulal, DeogratiasMzurikwao, Oluwagbenga Paul Idowu et al. "A new technique for the prediction of heart failure risk driven by hierarchical neighborhood component-based learning and adaptive multi- layer networks." *Future Generation Computer Systems* 110 (2020): 781-794.
[81]. Hemanth, D. J. "Data mining technique based critical disease prediction in medical field." *Intelligent Systems and Computer Technology* 37 (2020): 104.
[82]. Marimuthu, M., M. Abinaya, K. S. Hariesh, K. Madhankumar, and V. Pavithra. "A review on heart disease prediction using machine learning and data analytics approach." *International Journal of Computer Applications* 181, no. 18 (2018): 20-25.
[83]. Tarawneh, Monther, and OssamaEmbarak. "Hybrid approach for heart disease prediction using data mining techniques." In *International Conference on Emerging Internetworking, Data & Web Technologies*, pp. 447-454. Springer, Cham, 2019.
[84]. Rjeily, Carine Bou, Georges Badr, Amir Hajjarm El Hassani, and Emmanuel Andres. "Medical data mining for heart diseases and the future of sequential mining in medical field." In *Machine Learning Paradigms*, pp. 71-99. Springer, Cham, 2019.
[85]. Beyene, Chala, and Pooja Kamat. "Survey on prediction and analysis the occurrence of heart disease using data mining techniques." *International Journal of Pure and Applied Mathematics* 118, no. 8 (2018): 165-174.
[86]. Enriko, I. Ketut Agung. "Comparative study of heart disease diagnosis using top ten data mining classification algorithms." In *Proceedings of the 5th International Conference on Frontiers of Educational Technologies*, pp. 159-164. 2019.
[87]. Kumar, M. N ikhil, K. V. S. Koushik, and K. Deepak. "Prediction of heart diseases using data mining and machine learning algorithms and tools." *International Journal of Scientific Research in Computer Science, Engineering and Information Technology* 3, no. 3 (2018): 887-898.
[88]. Kurian, Ria A., and K. S. Lakshmi. "An ensemble classifier for the prediction of heart disease." *International Journal of Scientific Research in Computer Science* 3, no. 6 (2018): 25-31.
[89]. Kumar, B. Senthil. "A Survey on Data Mining Techniques for Prediction of Heart Diseases." *IOSR Journal of Engineering (IOSRJEN) www. iosrjen. org ISSN* 8 (2018): 22-27.
[90]. Ramalingam, V. V., AyantanDandapath, and M. Karthik Raja. "Heart disease prediction using machine learning techniques: a survey." *International Journal of Engineering &Technology* 7, no. 2.8 (2018): 684-687.
[91]. Ghorbani, Ramin, and RouzbehGhousi. "Predictive data mining approaches in medical diagnosis: A review of some diseases prediction." *International Journal of Data and Network Science* 3, no. 2 (2019): 47-70.
[92]. Bagavathy, S., V. Gomathy, S. Sheeba Rani, Monica Murugesan, K. Sujatha, and M. K. Bhuvana. "Early heart disease detection using data mining techniques with hadoop map reduce Early Heart Disease Detection Using Data Mining Techniques with Hadoop Map Reduce." *International Journal of Pure and Applied Mathematics* 119, no. 12 (2018): 1915-1920.
[93]. Rajdhan, Apurb, Avi Agarwal, Milan Sai, Dundigalla Ravi, and Poonam Ghuli. "Heart disease prediction using machine learning." *International Journal of Research and Technology* 9, no. 04 (2020): 659-662.
[94]. Kinge, Durga, and S. K. Gaikwad. "Survey on data mining techniques for disease prediction." *International Research Journal of Engineering and Technology (IRJET)* 5, no. 01 (2018): 630-636.
[95]. Dubey, Animesh Kumar, and Kavita Choudhary. "A systematic review and analysis of the heart disease prediction methodology." *International Journal of Advanced Computer Research* 8, no. 38 (2018): 240-256.
[96]. Mokashi, Ankita R., Madhuri N. Tambe, and Pooja T. Walke. "Review on Heart Disease Prediction using Ann and Classifier." *International Journal of Engineering Research and Technology (IJERT)* 4, no. 9: 966-968.
[97]. Aldallal, Ammar, and Amina Abdul Aziz Al-Moosa. "Using Data Mining Techniques to Predict Diabetes and Heart Diseases." In *2018 4th International Conference on Frontiers of Signal Processing (ICFSP)*, pp. 150-154. IEEE, 2018.
[98]. Singh, Navdeep, Punjab Ferozepur, and Sonika Jindal. "Heart disease prediction using classification and feature selection techniques." *International Journal of Advance Research, Ideas and Innovations in Technology* 4, no. 2 (2018).
[99]. Terrada, Oumaima, BouchaibCherradi, Abdelhadi Raihani, and Omar Bouattane. "Classification and Prediction of atherosclerosis diseases using machine learning algorithms." In *2019 5th International Conference on Optimization and Applications (ICOA)*, pp. 1-5. IEEE, 2019.
[100]. Garg, Ravi, Elissa Oh, Andrew Naidech, Konrad Kording, and Shyam Prabhakaran. "Automating ischemic stroke subtype classification using machine learning and natural language processing." *Journal of Stroke and Cerebrovascular Diseases* 28, no. 7 (2019): 2045-2051.