A Project Report

on

Cardiac Disorders Prediction with Logistic Regression Algorithm using Machine Learning and Artificial Intelligence.

Submitted in partial fulfilment of the
requirement for the award of the degree of

Bachelor of Technology in Computer Science and
Engineering

GALGOTIAS
UNIVERSITY
(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

Under the supervision of:
Ms. Urvashi Sugandh
Assistant Professor
Department of Computer Science and Engineering

Team - BT3130

Submitted By

19SCSE1010054 - Shivam Singh Vats

19SCSE1010066 - Arpit Garg

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA, INDIA
DECEMBER - 2021

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA

CANDIDATE'S DECLARATION

I/We hereby certify that the work which is being presented in the thesis/project/dissertation, entitled **"CARDIAC DISORDERS PREDICTION WITH LOGISTIC REGRESSION ALGORITHM USING MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE"** in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology -submitted in the School of Computing Science and Engineering of Galgotias University, Greater Noida, is an original work carried out during the period of month, Year to Month and Year, under the supervision of Ms Urvashi Sugandh, Assistant Prefessor, Department of Computer Science and Engineering/Computer Application and Information and Science, of School of Computing Science and Engineering , Galgotias University, Greater Noida.

The matter presented in the thesis/project/dissertation has not been submitted by me/us for the award of any other degree of this or any other places.

Shivam Singh Vats, 19SCSE1010054
Arpit Garg, 19SCSE1010066

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Ms Urvashi Sugandh

Assistant Professor

CERTIFICATE

The Final Thesis/Project/ Dissertation Viva-Voce examination of Shivam Singh Vats: 19SCSE1010054 and Arpit Garg: 19SCSE1010066 has been held on _________________ and his/ her work is recommended for the award of Degree in Bachelors of Technology in Computer Science and Engineering~~.~~

Signature of Examiner(s) **Signature of Supervisor(s)**

Signature of Project Coordinator **Signature of Dean**

Date: December, 2021

Place: Greater Noida

ABSTRACT

The cases of heart attack diseases are increasing day by day at a rapid rate. It is one of the most critical human diseases in the world and affects human life very badly. Accurate and on time diagnosis of heart disease is important for heart failure prevention and treatment. The diagnosis of heart disease through traditional medical history has been considered as not reliable in many aspects.

Echocardiogram examination is important for diagnosing cardiac septal defects. With the development of AI-based technology, an echocardiogram examination previously performed manually by cardiologists can be done automatically. As the human assisted manual categorization can result in false/wrong prediction and diagnosis, septal segmentation is one of the most important and difficult task in the field of medical image processing. Furthermore, it is a difficult process when there is a huge amount of data to assist. Automatic segmentation of cardiac septal defects can help a physician to make an initial diagnosis before referring a pediatric cardiologist for further treatment. The enhancement of machine learning and technology will aid cardiologists in diagnosing these defects without taking invasive steps. In previous studies, automatic object segmentation using Logistic Regression was one of the DL applications that have been developed for cardiac abnormalities. This paper fundamentally centers around which patient is bound to have a heart illness dependent on different clinical characteristics. We have designed a system for the prediction of heart disease. We have designed a heart disease expectation framework to foresee whether the patient is probably going to be determined to have a coronary illness or not utilizing the clinical history of the patient. In this study, we propose a Logistic Regression-based U-Net architecture to automatically segment the chamber of heart and classify the patients if they are having probability of heart disease or not. Cardiovascular disease generally refers to conditions that involve narrowed or blocked blood vessels that can lead to a heart attack, chest pain (angina) or stroke. Other heart conditions, such as those that affect your heart's muscle, valves or rhythm, also are considered forms of heart disease.

In the proposed research paper, first we will use Convolutional Neural Network(CNN) algorithm that use deep learning for image classification and recognition in order to detect defects. It will automatically detect without any human supervision. Logistic Regression Machine Learning (and eventually comparing them) a analysis process to predict the probability of a discrete outcome for any input variable will be used for classifying whether a person is suffering from heart disease or not, for the given used dataset by predicting the output in binary form of Yes or NO.

A very Helpful methodology was utilized to direct how the model can be utilized to work on the precision of expectation of Heart Attack in any person. The strength of the proposed model hushed up fulfilling and had the option to anticipate proof of having a heart disease in a specific person.

Table of Contents

2. INTRODUCTION

These days, the cardiac diseases are quite possibly the most basic problem identifying with human wellbeing. The treatment of cardiac defects has been stated in a study that has gotten most consideration in the clinical framework around the world. Cardiovascular disorders are one of the most main reasons for death around the world. On middle, 17.7 million deaths result from coronary diseases which counts for around 31% all through the world in 2016, as per World Health Organization (WHO) [1]. The cardiovascular cases number, as the focal point of this review, shows that 82% of the cases are from low and center nations, 17 million are under 70 years old and inclined to noninfectious sicknesses, 6.7 million are impacted by stroke, and 7.4 million individuals are experiencing coronary illness (WHO, 2016) [2]. In the US and other created nations, about portion of all passings are brought about by coronary disease; additionally, 33% surprisingly's demises overall are identified with heart diseases. Heart disease influences individuals' wellbeing as well as the economies and expenses of nations also. The most widely recognized cardiovascular issues are those of microvascular beginning, principally heart issues and stroke. Following quite a while of openness to undesirable ways of life, cardiovascular infection clinically introduces itself in beginning phases of life, just as at an advanced age. The really cardiovascular ailments incorporate overweight, diabetes, family ancestry, smoking, and elevated cholesterol.

Congenital heart disease (CHD) is the most common congenital anomaly in new-born babies. Anatomical abnormalities of the heart and blood vessels have even occurred since the first trimester intrauterine. There are many types of CHD, varying from mild to severe, with both frequent and rare cases. A Cardiac septal defect is one type of CHD that is marked by a hole in the atrial, ventricular, or both the atrial and ventricular septa, which correspond to atrial septal defect (ASD), ventricular septal defect (VSD), and atrioventricular septal defect (AVSD), respectively. Even though AVSD is not as common as ASD and VSD cases, generally, the symptoms are more severe and overdue for detection.

To identify cardiac septal defects, the physician performs auscultation using a stethoscope to listen for heart sounds and murmurs. Although the first heart sound is normal, the second heart sound is typical of a wide fixed split and a soft systolic ejection murmur is heard over the pulmonary area in the left upper sternal border misdiagnosis of is still common.

These days, computer-based diagnosis systems have been developed. Echocardiogram interpretation is done digitally, aided by a computer device (computer-aided diagnosis) using artificial intelligence (AI). With the development of AI-based technology, an echocardiogram examination for the detection of cardiac septal defects previously performed manually by cardiologists can be done automatically. An automatic echocardiogram examination can be used to assist physicians in early detection before referral to a cardiologist for further management.

Object segmentation is one of the DL applications that can be developed for cardiac septal defects.

This Project’s study and contributions are as follows:

- To design a CNN model for segmenting cardiac septal defect conditions of the heart images with high accuracy;
- To develop a CNN-based U-Net architecture for segmenting the contour regions of ASD, blockage, strokes, and normal condition; and
- To validate selected models with a V-Net architecture in terms of pixel accuracy, mean intersection union, mean accuracy, precision

To determine the heart disease incident probability, the specific issues which should be examined are those identified with the practices. Besides, patients will go through broad assessments, for example, circulatory strain, glucose, important bodily functions, chest torment, electrocardiograms, greatest pulse, and raised degrees of sugar, however the splendid side might be that fruitful treatment is practical assuming the sickness is effectively and early identified and expected, yet therapy for these cardiovascular patients is relying upon clinical examinations, the patient history, and the reactions to inquiries by the patient. These strategies (history investigation, actual assessment exploration, and clinical expert assesses) regularly cause mistaken analysis and mechanical disappointment other than postponing the determination tests. Moreover, it is additionally more costly and calculation serious, and it sets aside a great deal of effort for assessments to be completed.

When diagnosing different diseases utilizing a PC based framework, high-layered and heterogeneous information should be handled and broke down. High-layered information much of the time brings about model overfitting and builds the preparation season of the model. Highlight determination is a dimensionality decrease method used to lessen a bunch of extensive number of information elements to a reasonable set by removing immaterial and repetitive elements that altogether affect classifier execution. Various compelling component determination techniques have been grown as of late to moderate the scourge of dimensionality. Calculations for include determination can be delegated regulated, unaided, or semisupervised [3]. Directed component determination is the underlying and most common strategy for choosing highlights that utilization named information. Three methodologies are used in managed highlight determination strategies: channel, covering, and implanted methodologies.

Deciding the probability of having cardiac defect physically is difficult to rely upon as hazard factors. As of late, to settle troublesome issues, a scope of information mining procedures and AI methods are constructed. All things considered, further developed AI will help us to distinguish designs and their valuable information. While it has a few uses in the clinical field, AI is primarily used to figure the coronary illness. To analyze defects, numerous analysts have been keen on using AI since it limits indicative time and shows precision and adequacy. Utilizing AI strategies, actually, a few defects can be distinguished, however heart finding is the fundamental target of this article since coronary illness is the main source of death these days and since effective coronary illness analysis is profoundly useful in saving lives.

AI (ML) assumes to be a critical part in cardiac defects anticipating. It predicts whether or not the patient has a specific defect type dependent on an effective learning procedure. In this paper, we are using administered learning methods for foreseeing the beginning phase of coronary illness. Troupe calculations and a few calculations, for example, a k-closest neighbor (KNN), support vector machine (SVM), choice tree (DT), Naive Bayes (NB), and arbitrary backwoods (RF) are utilized to group whether individuals tried have a place with the class of heart disease or sound individuals. Moreover, two strategies for include extraction, straight discriminant investigation (LDA) and head part examination (PCA), are utilized to choose fundamental highlights from the dataset.

3. LITERATURE SURVEY

Ritu et al. [4] introduced a feature determine technique for recognizing mortality occasions in patients with coronary illness during treatment to track down the most basic elements Various AI techniques are used, including LDA, KNN, RF, SVM, DT, and GBC. Moreover, the disarray lattice, recipient working trademark bend, accuracy, review rate, and F1-score are likewise produced to approve the SFS calculation's outcomes. The test discoveries showed that the consecutive component determination method accomplishes a precision of 86.67% for the irregular timberland classifier.

Gao et al. [5] recommended a model utilizing approach of ensemble (supporting and stowing) with highlight extraction calculations (LDA and PCA) for anticipating coronary illness. The creators analyzed outfit strategies (stowing and helping) with five classifiers (SVM, KNN, RF, NB, and DT) on chose highlights from the Cleveland coronary illness dataset. The consequences of the analyses showed that the packing group learning strategy with DT and PCA highlight extraction acquired the most exceptional exhibition.

Takci [6] utilized twelve arrangement calculations from different classes and four component determination techniques for respiratory failure expectation. The models were assessed dependent on the precision of the model, the handling time, and the discoveries of the ROC investigation. The outcome shows that, without highlight determination, the greatest exactness esteem was 82.59%; it expanded to 84.81% with include choice. Model precision of 84.81% was gotten utilizing innocent Bayes and straight SVM. Also, the handling time was diminished from 359 to 187 milliseconds. The ReliefF calculation gives the best model precision among the four elective element choice strategies as indicated by the mean exactness esteem. In this manner, the creator referenced that when the proper mixes are thought of, highlight determination has a useful impact in coronary failure forecast research.

Latha et al. [7] proposed a model to anticipate cardiac diseases hazard utilizing outfit grouping strategies and element choice methods. The review discoveries showed that gathering approaches like sacking and helping are valuable in expanding the expectation exactness of feeble classifiers and perform well in anticipating coronary illness hazard. Group characterization worked on the precision of powerless classifiers by up to 7%. The exhibition was additionally expanded by adding an element choice execution, which brought about a significant expansion in expectation exactness.

KarenGárate-Escamila et al. [8] proposed a hybreed technique that uses ACP to reduce the dimensions by joining joining Chi-square and head part investigation (CHI-PCA) to

anticipate cardiac disease. Their review was directed on three diverse datasets: Hungarian,Cleveland, and Hungarian-Cleveland datasets assembled from the UCI Machine Learning Repository. The exhibition of the recommended procedure was assessed with five unique classifiers: arbitrary backwoods, slope helped tree, choice tree, multi-facet perceptron, and strategic relapse. Chi-square and head part investigation (CHI-PCA) utilizing irregular backwoods (RF) showed the most noteworthy precision, at 98.7% for the Cleveland dataset, 99.0% for the Hungarian dataset, and 99.4% for the Cleveland–Hungarian (CH) dataset, separately.

Spencer et al. [9] directed examinations on four as often as possible utilized.datasets utilizing four diverse element choice methods: head part investigation, Chi-squared testing, ReliefF, and balanced vulnerability. As indicated by the creators, the advantages of component choice contrast contingent upon the AI approach utilized for the cardiovascular datasets. For instance, quite possibly the most reliable model found had an exactness of 85.0%, an accuracy of 84.73%, and a review of 85.56% when Chi-squared element determination was joined with the BayesNet classifier.

Semen et al. [10] developed an indicative model for ongoing capital recognition using a dataset of 400 patients with 24 elements. Recursive component end (RFE) was utilized to choose the main elements. The k-closest neighbors (KNN), support vector machine (SVM), choice tree, and arbitrary backwoods grouping techniques were utilized in this work. All order strategies accomplished fantastic execution. The arbitrary woods technique beat any remaining calculations, accomplishing 100% exactness, review, accuracy, and F1-score on all measurements.

Almansour et al. [11] utilized AI techniques to aid the early location of ongoing contour region abnormalities by contrasting two classifiers, SVM and ANN, further developing their boundaries utilizing an arbitrary thorough hunt strategy. The 400-occurrence dataset from the UCI store is preprocessed, and highlights are picked utilizing relationship coefficients. The classifiers' presentation is thought about against different best elements (2, 3, 6, 12, and all) and the registering time needed to prepare the model. At long last, the twelve best highlights are utilized to anticipate renal sickness utilizing SVM and ANN since ANN outperforms SVM with high exactness.

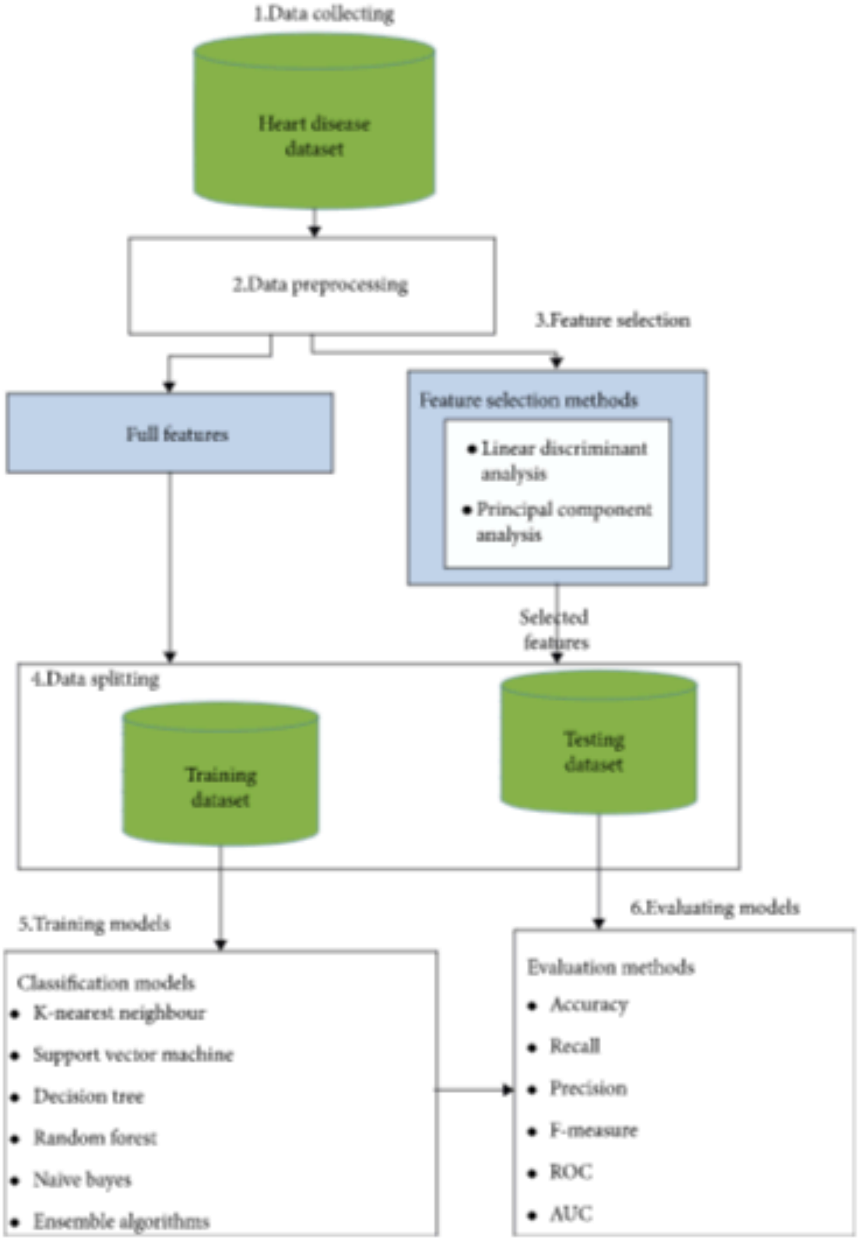
There are numerous literatures related to coronary defects that analyze utilizing information mining and AI methods. Reddyet al. utilized RF, SVM, NB, NN, and

KNN with numerous element choice like connection grid, recursive component disposal (RFE), and learning vector quantization (LVQ) model to characterize the cardiovascular infection into ordinary or unusual. The outcomes show that RF achieved the ideal exhibition. Atallah and Al-Mousa used stochastic slope plummet (SGD), KNN, RF, calculated relapse (LR), and casting a ballot gathering figuring out how to foresee cardiovascular illnesses. The democratic outfit learning model has accomplished the best exactness of 90%.

Current literature have effectively shown that it is feasible to recognize standard echocardiogram sees from unlabeled datasets.5,6,24 Previous works have utilized CNNs prepared on pictures and recordings from echocardiography to perform division to distinguish cardiovascular designs and determine heart work. In this review, we stretch out past investigations to show that EchoNet, our profound learning model utilizing echocardiography pictures, can dependably distinguish neighborhood heart designs and life structures, gauge volumetric estimations and measurements of heart work, and foresee fundamental human aggregates that change cardiovascular danger. Also, we show the primary use of translation systems to see profound taking in models from echocardiogram pictures. Human-recognizable elements, for example, the presence of pacemaker and defibrillator drives, left ventricular hypertrophy, and unusual left atrial chamber size distinguished by our CNN were approved utilizing translation structures to feature the most significant areas of interest. Supposedly, we foster the main profound leaming model that can straightforwardly foresee age, sex, weight, and range from echocardiogram pictures and use translation strategies to see how the model predicts these foundational aggregates hard for human translators.

3. PROPOSED WORKING

The objective of the proposed working procedure is to utilize troupe strategies to work on the exhibition of foreseeing Cardiac defects. Figure below depicts the design of the proposed framework. It is organized into six phases, including information assortment, information preprocessing, highlight determination, information parting, preparing models, and assessing models.



The heart disease dataset which is utilized for training and evaluating models consists of 1025 records, 12 features, and one target column. The target column includes two classes: 1 indicates heart diseases, and 0 indicates nonheart disease.

Feature ID	Feature name	Feature type	Description	Values
F1	AGE	Numerical	Age in years	28–77; Mean: 51.9
F2	SEX	Nominal	Gender (1: male; 0: female)	1: (206); 0: (97)
F3	CP	Nominal	Chest pain type (1: typical angina; 2: atypical angina; 3: nonanginal pain; 4: asymptomatic)	1: (23) 2: (50) 3: (86) 4: (144)
F4	TRESTBPS	Numerical	Resting blood pressure (in mmHg on admission to the hospital)	94–200; mean: 131.6
F5	CHOL	Numerical	Serum cholesterol (in mg/dl)	126–564; mean: 246.6

F6	FBS	Nominal	Fasting blood sugar >120 mg/dl (1: true; 0: false)	1:(45); 0: (258)

F7	RESTECG	Nominal	Resting electrocardiographic results (2: showing probable or definite left ventricular hypertrophy by Estes' criteria, month of exercise ECG reading; 1: having ST-T wave abnormality; 0: normal)	2: (148);
				1: (4);
				0: (151)

F8	THALACH	Numerical	Maximum heart rate achieved	71–202; mean: 149.6
F9	EXANG	Nominal	Exercise-induced angina (1: yes; 0: no)	1: (99); 0: (204)
F10	OLDPEAK	Numerical	ST depression induced by exercise relative to rest	0–6.2; mean: 1.03

F11	SLOPE	Nominal	The slope of the peak exercise ST segment (3: downsloping; 2: flat; 1: upsloping)	3: (21); 2: (140); 1: (142)
F12	CA	Nominal	Number of major vessels (0–3)	3: (24); 2: (38); 1: (65); 0: (176)

3.1 Data Preprocessing

Preprocessing the dataset is essential for the viable portrayal of the of data quality. The dataset has been preprocessed utilizing strategies, for example, removing missing qualities from highlights, StandardScaler (SS), and MinMaxScaler. Missng value hndling is an information preprocessing method used to make a smooth dataset. In this way, it was begun by deciding whether the dataset contains any missing qualities [2, 5]. The missing qualities can be taken care of in different ways, including absolutely disregarding the missing worth, supplanting it with any numeric worth, supplanting it with the regularly happening (mode) an incentive for that component, or supplanting it with the mean worth of the property. The coronary illness dataset utilized in this article comprises of six missing qualities. The missing qualities are supplanted with the mean or mode upsides of the elements.

The features are scaled to be in the interval [0, 1]. It is worth noting that missing values are deleted from the dataset.

3.2 Data Splitting

To train our Machine, we see the section in the information column, then, at that point, we partition the column index into two little datasets. Preparing set to prepare the calculation is the Test-set to test it. Fig. belowexplain show we did to apply the ML algorithms, Where we first divide our dataset in two sections as referenced previously, then, at that point, in Training Set we partition the data once more, for preparing and approval (This subsequent advance is done automatically).
To test our model with respect to its predictive precision it appears to be very instinctive to divide information into a preparation segment and a test segment, so the model can be prepared on one dataset, yet tried on an alternate, new information segment. Naturally, the concept of train, validation, and test influences the way you should process your data as you are getting ready for training and deployment of your computer vision model.

- **Training Dataset**: The sample of data used to fit the model.
- **Validation Dataset:** The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters. The evaluation becomes more biased as skill on the validation dataset is incorporated into the model configuration.
- **Test Dataset:** The sample of data used to provide an unbiased evaluation of a final model fit on the training dataset.

What percent of Data is obtained after splitting

70% train, 15% val, 15% test
80% train, 10% val, 10% test
60% train, 20% val, 20% test

Preprocessing Steps: These are image transformations that are used to standardize your dataset across all three splits. Examples include static cropping your images, or gray scaling them. All preprocessing steps are applied to train, validation, and test.

Image Augmentation: These are used to increase the size of your training set by making slight alterations to your training images. These occur only to the training set and should not be used during evaluation procedures. For evaluation, you want to use the ground truth images, residing in the validation and test sets.

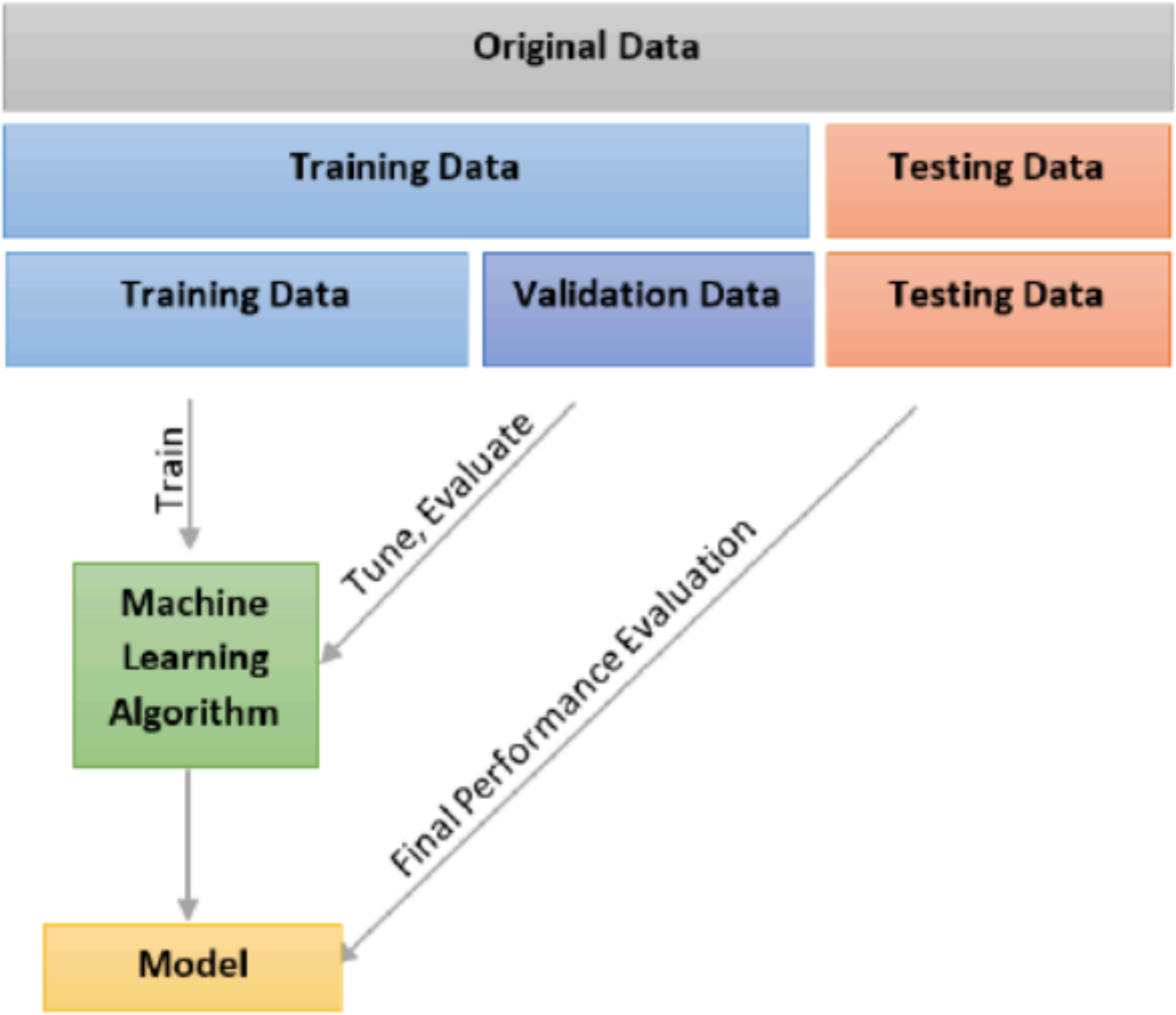
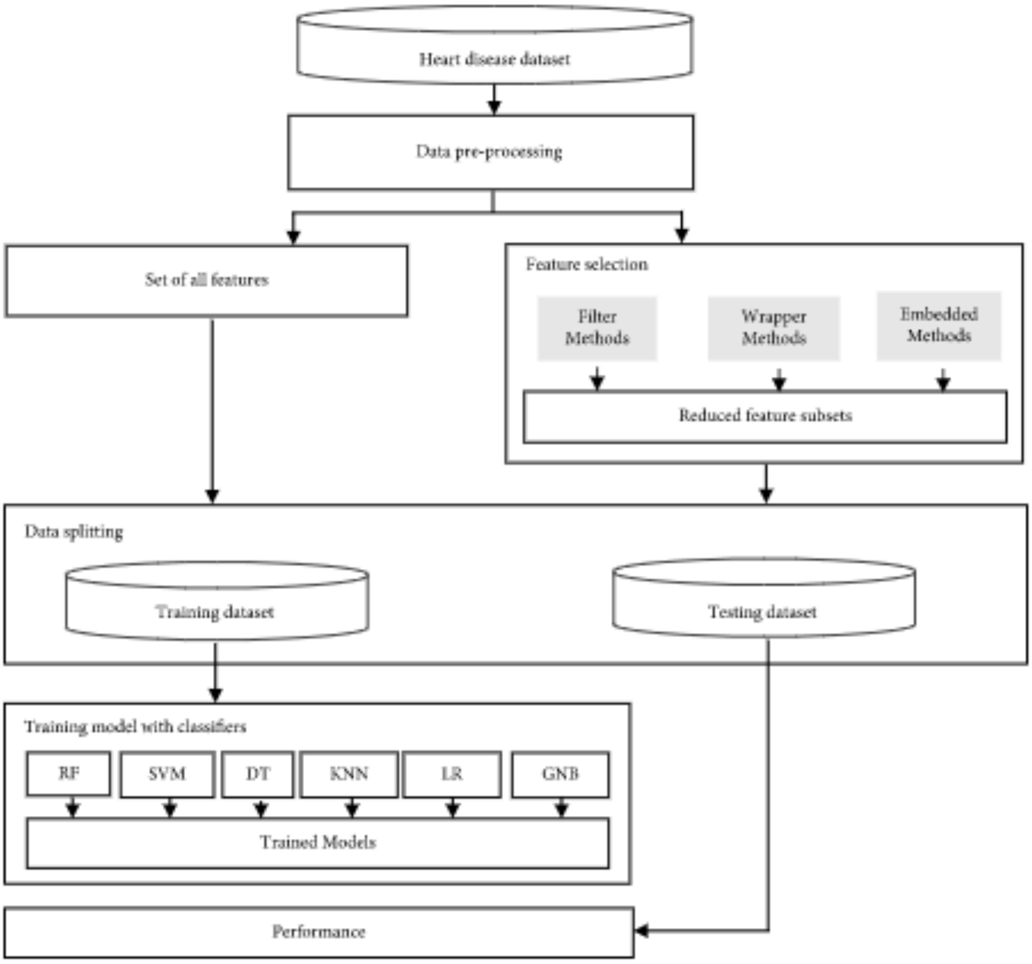


Figure showing the Splitting of Data

4. SYSTEM FRAMEWORK

The proposed study focuses on increasing classification accuracy by decreasing the number of features in a dataset of cardiac disease. The framework for classifying cardiac diseases is depicted in Figure 1. The key components of the framework include data collection, data preprocessing, feature selection, data splitting, model training with classifiers, and model evaluation. The following sections describe the building blocks of the suggested framework.



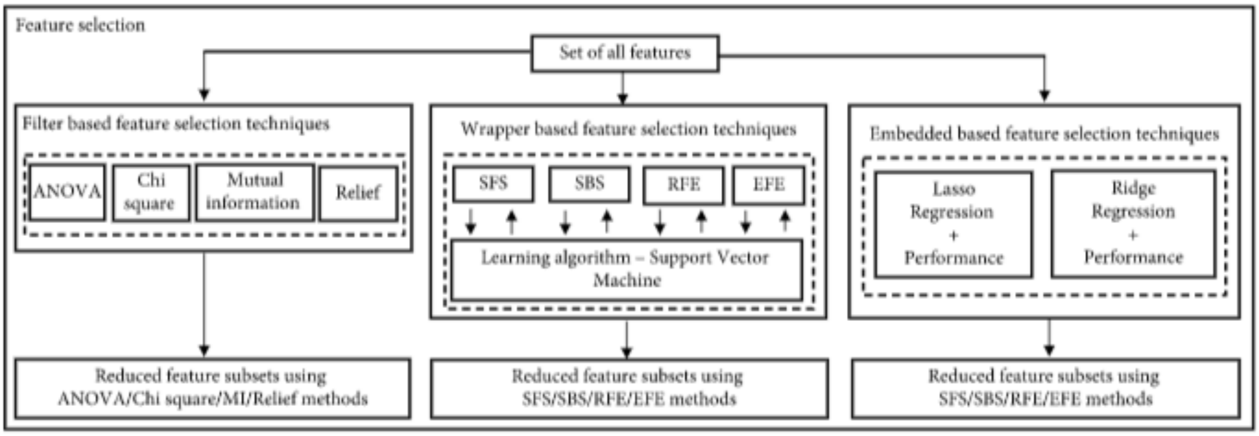
System framework

5. FEATURE SELECTION

Feature selection methods are expected to decrease the quantity of input dimensions to those that are accepted to be generally helpful to a model to predict the objective variable.

Some prescient displaying issues have an enormous number of factors that can slow the working and efficiency and preparing of models and require a lot of framework memory. Furthermore, the presentation of certain models can debase when including input factors that are not applicable to the objective variable.

In this stage, tests were led with and without feature selection to evaluate the impact of feature selection. The component choice plans to recognize the main elements of cardiovascular infections. Besides, highlight determination assists with developing a more precise model by disposing of or underrepresenting the less important elements, limiting preparing time and improving learning execution [6]. The conduct of a few element choice methodologies under the three significant classes (channel, covering, and inserted) is surveyed in this examination. As outlined in Figure 2, highlight choice methods having a place with three classes were independently applied to beginning datasets. Highlight choice methods start by making a subset, yet that subset age relies upon the sort of approach. Figure 2 shows the course of element choice followed by every one of the three classifications to recognize ideal component subsets.


Feature selection
Set of all features
Filter based feature selection techniques
ANOVA
Chi square
Mutual information
Relief
Wrapper based feature selection techniques
SFS
SBS
RFE
EFE
Learning algorithm – Support Vector Machine
Embedded based feature selection techniques
Lasso Regression + Performance
Ridge Regression + Performance
Reduced feature subsets using ANOVA/Chi square/MI/Relief methods
Reduced feature subsets using SFS/SBS/RFE/EFE methods
Reduced feature subsets using SFS/SBS/RFE/EFE methods

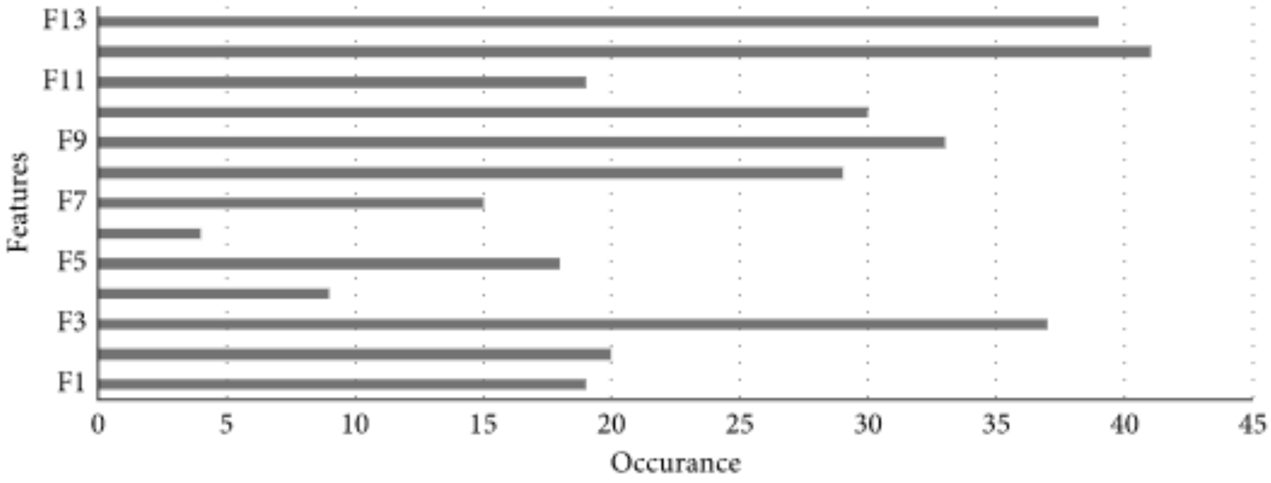
Feature selection framework of system

- **Feature Selection**: Select a subset of input features from the dataset.
    - **Unsupervised**: Do not use the target variable (e.g. remove redundant variables).
        - Correlation
    - **Supervised**: Use the target variable (e.g. remove irrelevant variables).
        - **Wrapper**: Search for well-performing subsets of features.
        - **Filter**: Select subsets of features based on their relationship with the target.
            - Statistical Methods
            - Feature Importance Methods
        - **Intrinsic**: Algorithms that perform automatic feature selection during training.
            - Decision Trees

5.1 Feature Selection Using Filter Methods

Filter-based selection techniques utilize statistical methods to determine the dependence or association among independent features (input attributes) and the dependent feature (target attribute).

In filter method, without depending on any learning algorithms, features are assessed based on the general properties of the data as represented in Figure 3. On the other hand, characteristics are chosen based on the output results/scores generated by the various statistical methods used to validate them.



It primarily quantifies the properties of features using distinct types of measurement criteria, including information, distance, dependence, consistency, similarity, and statistical measures. These attributes are compared for their interdependence with the output attribute to choose the appropriate subset.

- Chi-Square- The Chi-square test follows the univariate statistical approach in feature selection to identify the dependence of each feature (or correlation) with the target variable in a dataset [8]. Here, remove the features independent of the target variable and other dependent features considered highly important to predict the target variable.

- Mutual Information- The mutual information (MI) between two random variables is a metric measuring their interdependence or measures the reduction in uncertainty. The actions necessary to complete this process are listed below.

- ReliefF Algorithm- The ReliefF is a supervised feature learning approach to weigh each feature, demonstrating its importance to the target variable used for binary and multiclass problem domains. The steps required to accomplish this procedure are summarized below.

- ANOVA Method- ANOVA (Analysis of Variance) is a statistical test that formulates the ratio between variances, such as the variance from two different samples. Furthermore, ANOVA is beneficial in classification; the data come up with numeric input and categorical target variables [11, 19]. The steps necessary to complete this process are listed below.

5.2 Feature Selection Using Wrapper Methods

This method mainly utilizes a search strategy to evaluate the variable subsets of independent features S'⊆S by feeding S′ to the chosen learning algorithm and then assessing the performance of a learning algorithm. When the cardinality of features in a dataset comes up with N, the 2N subsets are possible, and the steps are done iteratively before the required suboptimal subsets are found. Thus, the value of the cardinality is very high, and the number of subsets is exponential. The searching method can identify various types of techniques and in this experiment mainly used four types of techniques: forward feature selection, backward feature elimination, recursive feature elimination, and exhaustive feature selection described below:

- Forward Feature Selection (FFS)- Forward feature selection is an iterative procedure that begins with the feature having the highest performance versus the target feature. When selecting the feature subset, with support vector machine used as a learning algorithm, stratification is required to guarantee that each class is well represented [4]. Each subset is evaluated independently, and the creation of subsets is determined by the search method. The following are the steps required to complete the feature selection procedure.

- Backward Feature Elimination (BFE)- The approach of backward elimination is the complete opposite of the method of forward feature selection [4]. The following stages are followed to finish the feature selection procedure.

- Recursive Feature Elimination (RFE)- Recursive feature elimination uses an approach to iteratively train the model, removing the least significant feature using the algorithm's weights as a criterion for each iteration. Recursive feature elimination aims to choose features by examining smaller and smaller sets of features in a recursive fashion.

- Exhaustive Feature Selection (EFS)- When determining the optimum feature subset, the exhaustive feature selection approach used brute-force strategies. The performance of the machine learning algorithm is assessed against all potential feature combinations in the dataset. The optimum feature subset is determined by the one that produces the best performance. The exhaustive search algorithm is the greediest algorithm of the wrapper techniques since it attempts all possible combinations of features and chooses the best.

5.3 Feature Selection Using Embedded Methods

Embedded techniques include integrating the feature selection process into the machine learning algorithm's development. This is a hybrid technique that combines the filter and wrapper methods. Here, the algorithms include their feature selection strategy. These contribute to the generation of the optimal subset and offer it to the training model. In other words, they pick features during model training, which is why they are referred to as embedded techniques. These techniques need less computation than wrapper methods.

- Lasso Regression- The abbreviation “Lasso” refers to the Least Absolute Selection and Shrinkage Operator. In Lasso, the penalty is defined as the total of the absolute values of the coefficients, which is L1. Thus, the primary goal of Lasso is to reduce an absolute value (L1 penalty) to zero rather than to use the sum of squares (L2 penalty)

- Ridge Regression- Ridge regression employs L2, with the penalty equal to the sum of the squares of the coefficients of the variables. This form of regression aids in the reduction of variation caused by variable multicollinearity. It contributes to the removal of variation caused by nonlinear correlations between two independent variables. The following steps describe how to use Ridge to choose the optimal feature subset.

Pre-processing

The pre-processing of video for segmentation consists of four main steps;

(i) video to image framing. The type of file.avi and length is 5s. cv2.VideoCapture (); is used. The video was then read frame by frame, where the frame will be stored in frame storage using cv2.imwrite () code to create a ground truth of infant images.

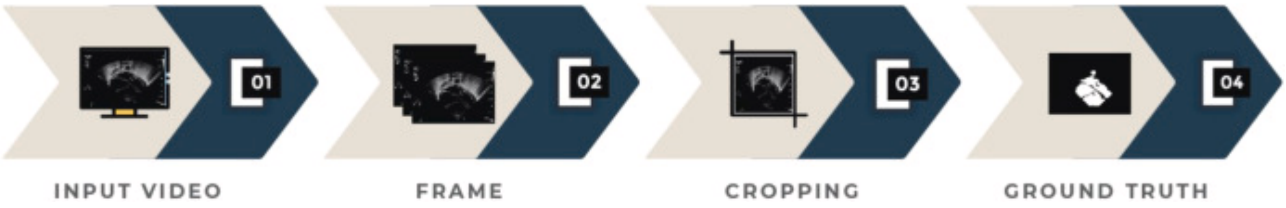
(ii) the data filtering process was performed with a closed valve case;

(iii) cropped all infant images from the frame based on an 800 × 600 pixel;

(iv) and annotated the labels of infant images with a data annotation tool (Adobe Photoshop). The label consisted of a heart chamber. If there were only the chambers of the heart, it was identified as normal. The output of labels was saved in image thresholding.

For videos that have been obtained previously, the next step is to convert videos into frames or images. From the raw video data inserted into the Python library with OpenCV, the video will be converted into many frames. The data is recorded in the video in the.avi format and then converted into frames with the.jpg format. the video was converted into frames of AVSD.

The results obtained in the process of converting the video to the frame will produce many frames based on the output obtained by the Python library. The data frame results obtained have a size of 800 × 600 pixels, and there is still much unnecessary information in the data frame. Thus, the next stage is to cut the image frame that has been performed before. This stage is performed the same as in the process of converting the video to the frame using library Python software. In the process of cropping the frame with the Python library, the crop range is adjusted to the right, left, top and bottom to remove unnecessary information.



6. MACHINE LEARNING ALGORITHMS

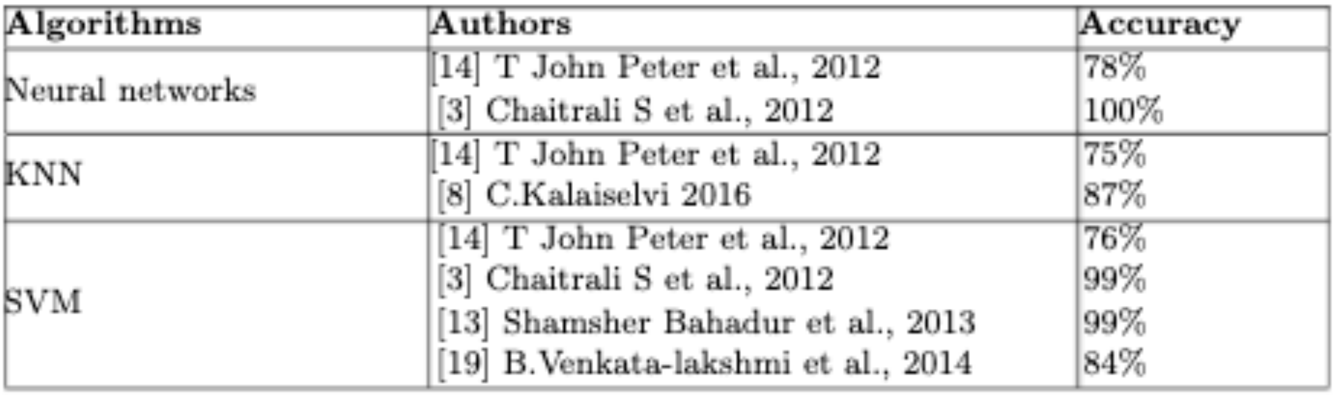
An efficient Machine Learning algorithm gives more accuracy. The prediction of heart patients is very critical, because a simple mistake can lead to death of a human being. This section consists of evaluating and selecting the most frequently used algorithms with high accuracy. As in the first section, we have summarized the most recent articles. We start with [6], The authors have imple-ented several learning machine algorithms, Logistic Regression has given an accuracy of 93%, Random Forest 92% and Gaussian Na¨ıve Bayes 90%, we gave notice that the results are close with simple progression of Logistic Regression.

The authors of [1], tested the diagnosis of heart patients by applying two techniques: genetic algorithms and the KNN algorithm. The results gave satisfaction with the KNN algorithm. The weak point of this article is that the authors did not mention their results. Same remark for [5], The authors have used artificial neural networks now unfortunately, they have not presented their results.

Priyanka.N et al [15], made a comparison between two algorithms (Na¨ıve bayes and KNN), the results show that KNN gives much more accuracy than na¨ıve bayes), it reaches 85% accuracy. The authors of [17], conducted a study on the prediction of heart disease, using only dataset, the algorithms used with their accuracy of SVM (99.3%), Neural Networks (91.1%), KNN (87.2%) and decision trees (82.3%). [18] The authors conducted a survey on the most used Machine Learning techniques that give more precision, Table 2 illustrates the algorithms cited in the articles with their accuracy.

At the end of the analysis of these papers, we decided to choose the important attributes of the three classes[6], and the algorithms to implement are: Neural Networks (NN), KNN and SVM, to test our dataset which contains information on Algerian patients.

Most frequently cited algorithms are:-



Algorithms	Authors	Accuracy
Neural networks	[14] T John Peter et al., 2012	78%
	[3] Chaitrali S et al., 2012	100%
KNN	[14] T John Peter et al., 2012	75%
	[8] C.Kalaiselvi 2016	87%
SVM	[14] T John Peter et al., 2012	76%
	[3] Chaitrali S et al., 2012	99%
	[13] Shamsher Bahadur et al., 2013	99%
	[19] B.Venkata-lakshmi et al., 2014	84%

Training Models

Classification is a method for labeling a particular set of data with different classes, which is carried out on structured and unstructured data. The aim of classification predictive modeling is to approximate mapping between input variables and discrete output variables. There are several classification algorithms available, but one that is better than another cannot be concluded. This relies on the domain of the problem and the nature of the dataset available.

Different types of machine learning algorithms: KNN, DT, RF, and NB are applied to classify heart disease. Also, two types of ensemble techniques: boosting and bagging are applied to classify heart disease:

6.1 K-Nearest Neighbor (KNN)

- *K*-nearest neighbor is a technique of supervised learning used to recognize the patterns in the dataset and determine the belonging class based on the points of existence with the nearest neighbors.
- It assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.
- Stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.

It determines the class of instance by considering the distance from the nearest neighbor and consistency.
KNN uses different distance functions such as Manhattan, Euclidean, Hamming, and Minkowski when identifying the neighbor's points. This method is called a memory dependent approach because all the instance points are stored inside the memory. The weight is allocated for each point depending on the distance from each instance to improve the algorithm's performance.

The following equation is used to calculate the Euclidean distance:

$$D = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}.$$

○ The K-NN working can be explained on the basis of the below algorithm:

- **Step-1:** Select the number K of the neighbors
- **Step-2:** Calculate the Euclidean distance of **K number of neighbors**
- **Step-3:** Take the K nearest neighbors as per the calculated Euclidean distance.
- **Step-4:** Among these k neighbors, count the number of the data points in each category.
- **Step-5:** Assign the new data points to that category for which the number of the neighbor is maximum.
- **Step-6:** Our model is ready.

6.2 Decision Tree (DT)

A decision tree algorithm can be categorized under supervised learning, which is identical to a tree of nodes and edges ideal for classification problems. Every node in the tree represents the class of the problem, and every edge indicates the choice made based on the evaluated results. This classifier can be considered a predictive machine learning model that displays the correlation between dataset values and features. Each division in the decision tree indicates the potential value for a certain category. Dependent on entropy measurements of the dataset attributes, the nodes are identified. The maximum entropy value attribute is known to be the root node. C4.5, Iterative Dichotomiser 3(ID3), and Classification and Regression Trees (CART) are widely used decision tree models.

Decision Tree Terminologies

Root Node: Root node is from where the decision tree starts. It represents the entire dataset, which further gets divided into two or more homogeneous sets.

Leaf Node: Leaf nodes are the final output node, and the tree cannot be segregated further after getting a leaf node.

Splitting: Splitting is the process of dividing the decision node/root node into sub-nodes according to the given conditions.

Branch/Sub Tree: A tree formed by splitting the tree.

Pruning: Pruning is the process of removing the unwanted branches from the tree.

Parent/Child node: The root node of the tree is called the parent node, and other nodes are called the child nodes.

Steps how decision Tree work

- **Step-1:** Begin the tree with the root node, says S, which contains the complete dataset.
- **Step-2:** Find the best attribute in the dataset using **Attribute Selection Measure (ASM).**
- **Step-3:** Divide the S into subsets that contains possible values for the best attributes.
- **Step-4:** Generate the decision tree node, which contains the best attribute.
- **Step-5:** Recursively make new decision trees using the subsets of the dataset created in step -3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.

6.3 Random Forest (RF)

One of the most widely used, technologically advanced, supervised ensemble classification methods is called random forest. It generates a large number of trees during the training stage and builds a forest of decision trees on multiple data subsets. At the testing stage, each tree in the forest assigns each piece of data a class label. When each tree predicts a class label, the ultimate determination for each set of test data is made by a majority vote. Whichever class label receives the greatest number of votes is assumed to be the right label for the test data to enhance its prediction accuracy. This procedure is performed for each piece of data in the dataset

- Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.
- The greater number of trees in the Random Forest leads to higher accuracy and prevents the problem of overfitting.

Random forest= base learner (DT)+ bagging+ feature bagging + aggregation

General Representation of Random Forest

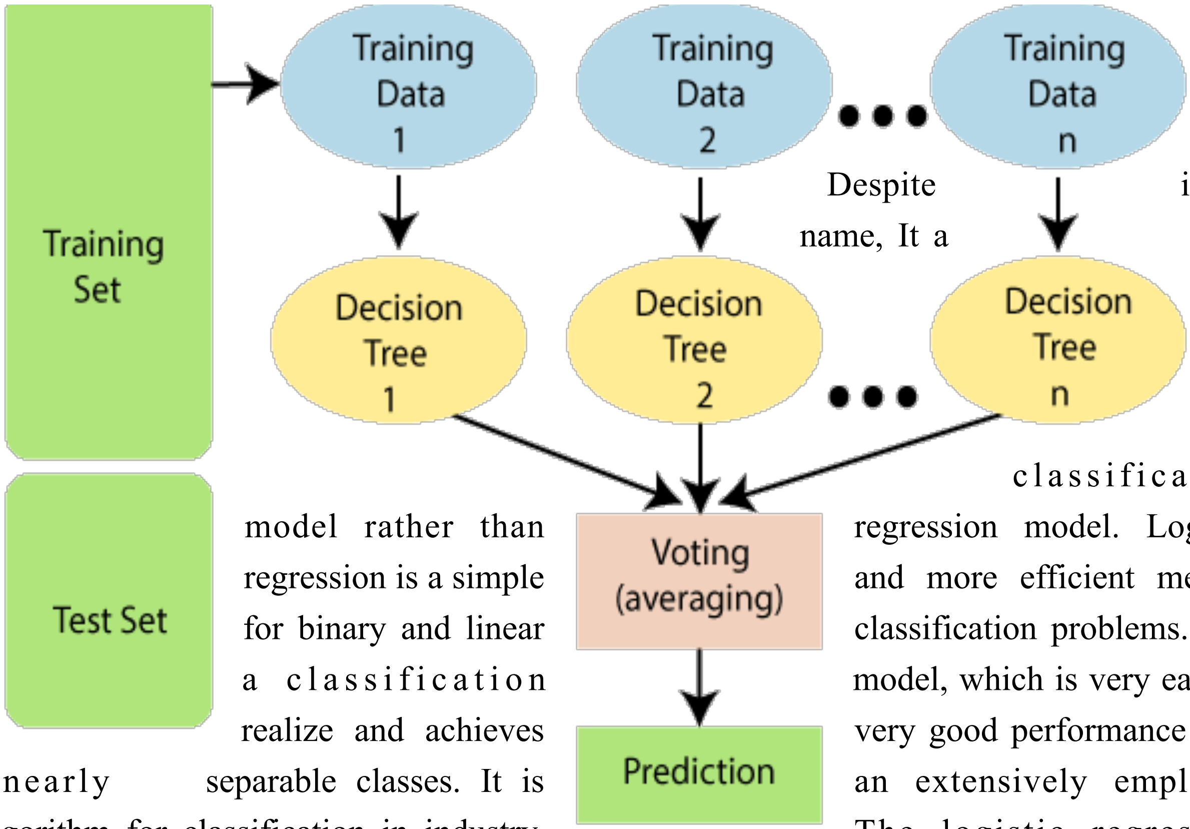
6.4 Logistic Regression

Logistic regression is a process of modeling the probability of a discrete outcome given an input variable. The most common logistic regression models a binary outcome; something that can take two values such as true/false, yes/no, and so on. Multinomial logistic regression can model scenarios where there are more than two possible discrete outcomes.

Logistic regression is a useful analysis method for classification problems, where you are trying to determine if a new sample fits best into a category. As aspects of cyber security are classification problems, such as attack detection, logistic regression is a useful analytic technique.

$$\textit{Logistic function} = \frac{1}{1+e^{-x}}$$

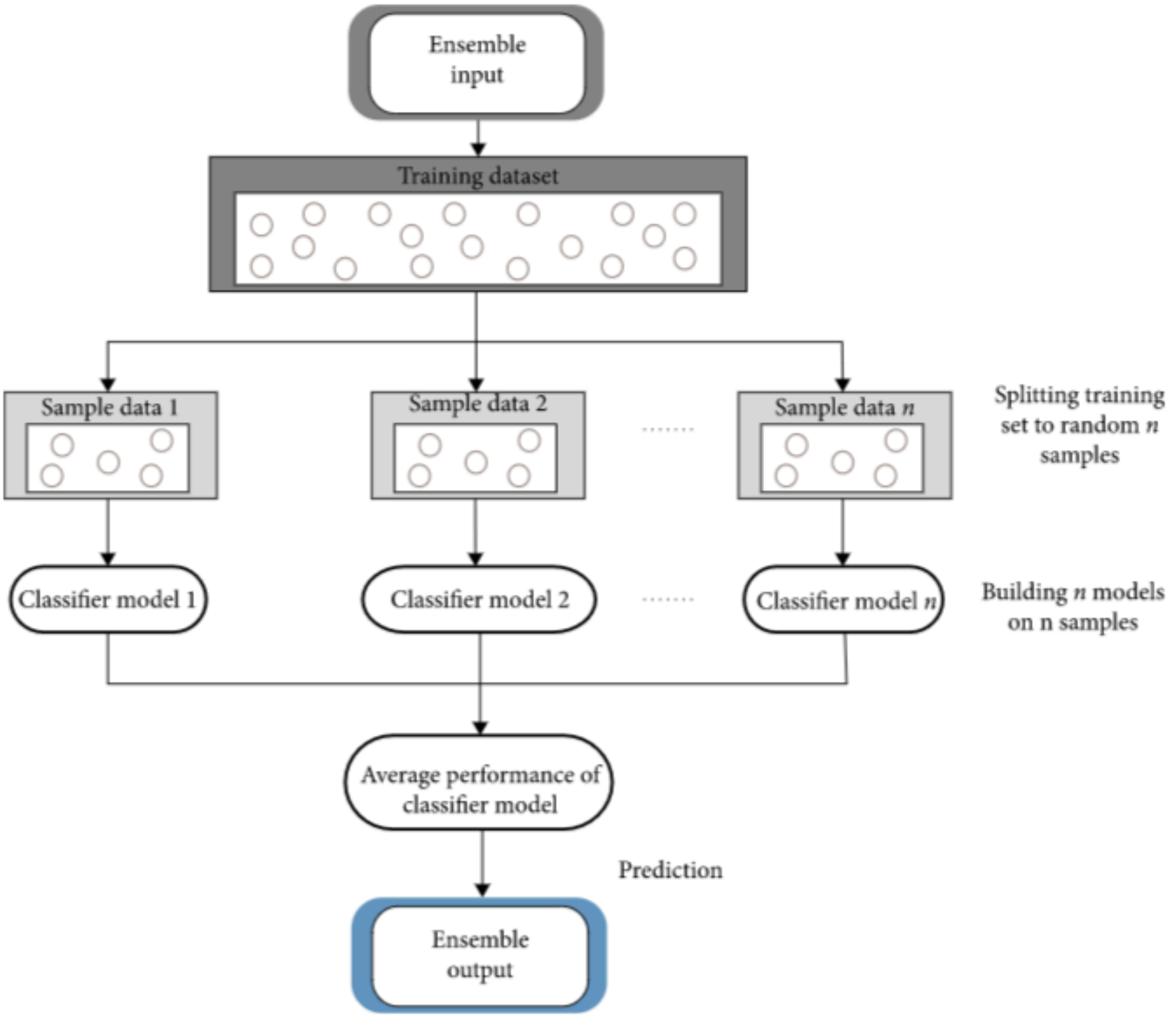
- Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, **it gives the probabilistic values which lie between 0 and 1**.
- Logistic Regression is much similar to the Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas **Logistic regression is used for solving the classification problems**.



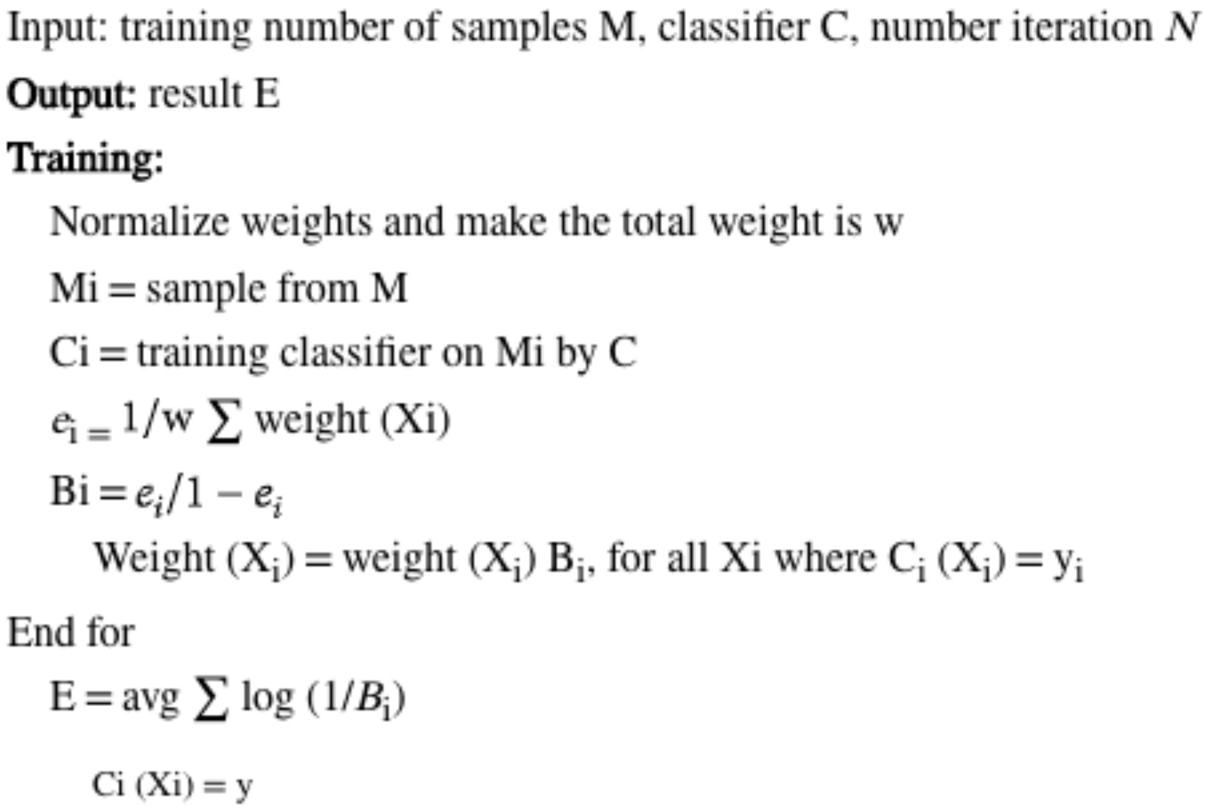
Despite its name, It a classification model rather than regression model. Logistic regression is a simple and more efficient method for binary and linear classification problems. It is a classification model, which is very easy to realize and achieves very good performance with linearly separable classes. It is an extensively employed algorithm for classification in industry. The logistic regression model, like the Adaline and perceptron, is a statistical method for binary classification that can be generalized to multiclass classification. Scikit-learn has a highly optimized version of logistic regression implementation, which supports multiclass classification task.

7. ENSEMBLE TECHNIQUES

Ensemble techniques are methods that can be utilized to enhance the performance of a classifier. It is an effective classification method that combines a weak classifier with a strong classifier to improve the weak learner's efficiency [34]. The ensemble technique is used in the proposed technique to enhance the accuracy of various algorithms for diagnosing heart disease. Compared to an individual classification, the purpose of combining multiple algorithms is to obtain better performance. Figure 2 explains how the ensemble approach is utilized to enhance heart disease diagnosis.



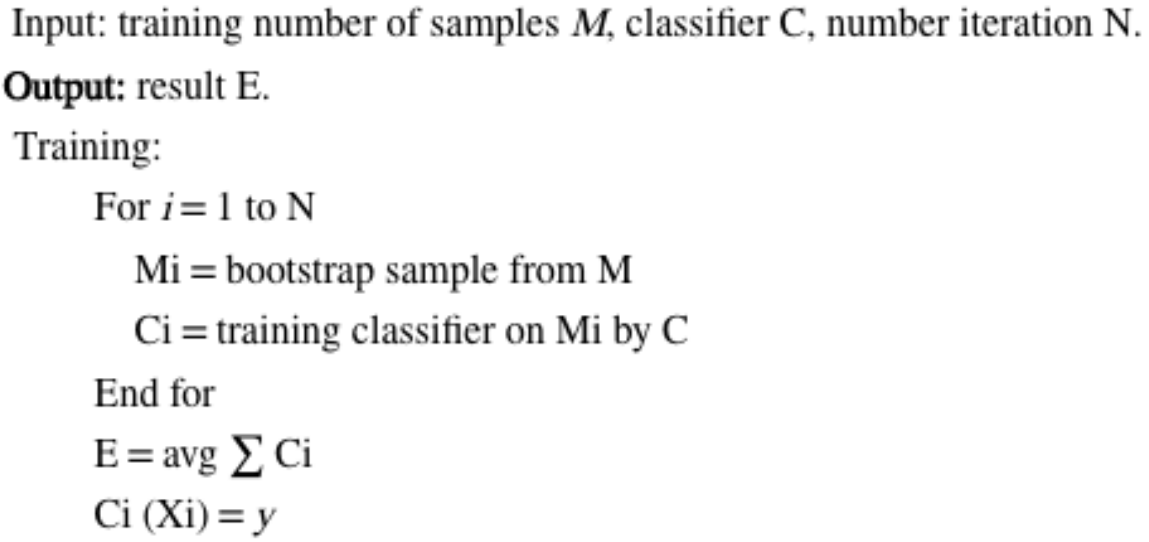
7.1 **Boosting:-** Boosting means producing a model sequence that aims to correct the errors that have arisen in the models. The dataset is split into different subsets in detail [35]. The classification algorithm is then trained on a sample to create a series of average efficiency models as shown in pseudocode of boost algorithm, where B is the number of base hypotheses and e is exp 1/e = 0.368. Consequently, based on the previous model's elements not properly classified, new samples are produced. Then, by combining the weak models, the ensemble method increases its efficiency. The pseudocode for boosting is provided in Algorithm below-



```
Input: training number of samples M, classifier C, number iteration $N$
Output: result E
Training:
   Normalize weights and make the total weight is w
   Mi = sample from M
   Ci = training classifier on Mi by C
   $e_i = 1/w \sum$ weight (Xi)
   Bi = $e_i/1 - e_i$
      Weight ($X_i$) = weight ($X_i$) $B_i$, for all Xi where $C_i$ ($X_i$) = $y_i$
End for
   E = avg $\sum \log (1/B_i)$
      Ci (Xi) = y
```

The pseudocode of boost algorithm.

7.2 **Bagging:-** It refers to taking a replacement training set with multiple subsets and training a model for each subset [35]. The average of the forecast values of the submodels together are as stated by the final performance forecast. A voting procedure for each classification model is then performed as shown in pseudocode of bagging algorithm. Consequently, the classification outcome is determined based on the majority of the average values. The pseudocode for bagging is provided in Algorithm below:



```
Input: training number of samples $M$, classifier C, number iteration N.
Output: result E.
Training:
    For $i = 1$ to N
        Mi = bootstrap sample from M
        Ci = training classifier on Mi by C
    End for
    $E = avg \sum Ci$
    $Ci (Xi) = y$
```

The pseudocode of boost algorithm.

8. EVALUATING MODELS

Various assessment metrics are applied to determine the efficiency of the feature selection techniques in this study, namely, precision, F1-score, accuracy, and recall, with the help of elements in the confusion matrix. The confusion matrix allows researchers to identify the performance rate of the result in classification problems depending on four main factors: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). The accuracy will be shown by the model if it can raise proper identification of high risk level of getting heart disease. Accuracy can be calculated by

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}.$$

Precision, calculated as a summation of correctly identified instances, means that the proportion of patients diagnosed with heart disease actually had risk. Precision is given by

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}.$$

Recall is the measure which indicates that the algorithm diagnosed the proportion of patients that actually had heart disease as having disease and is defined by

$$\text{recall or sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}.$$

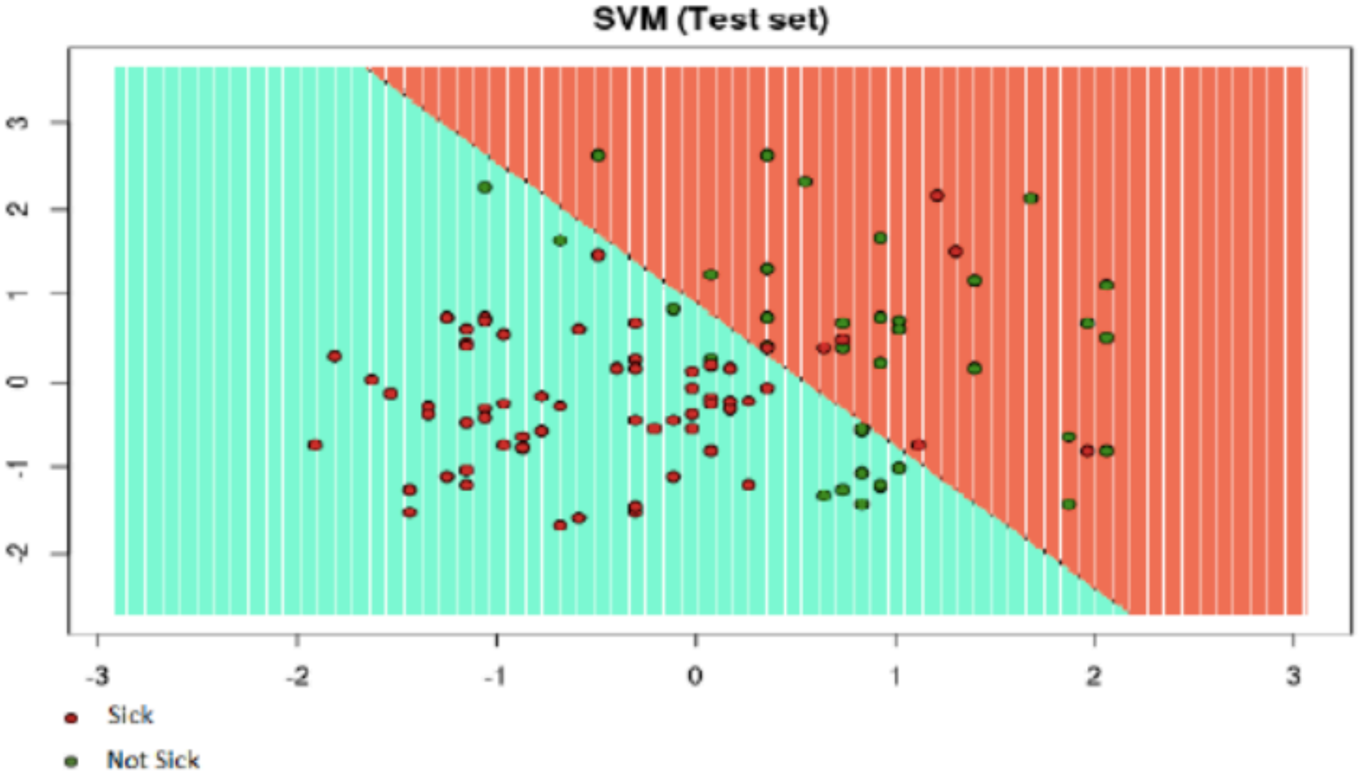
F-score computes the accuracy of the test, and the formula used to evaluate it is

$$\text{F} - \text{measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}.$$

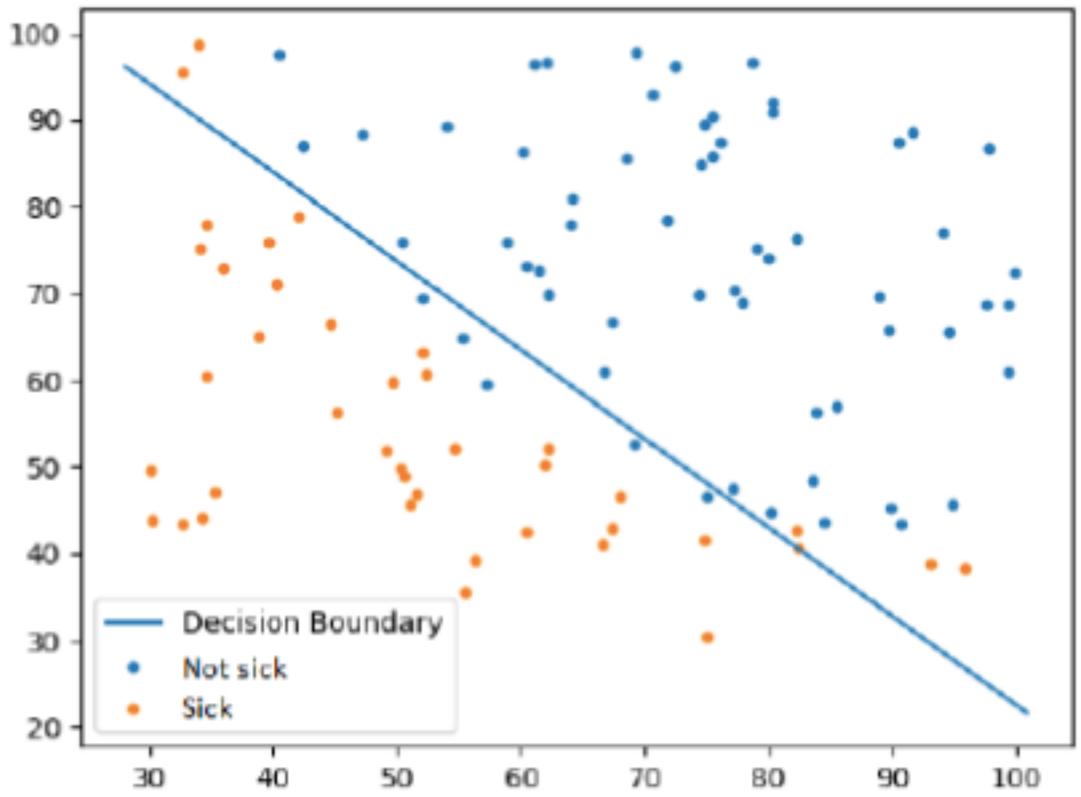
Modelling

We apply the selected algorithms (neural networks, SVM, KNN) and analyze the obtained results.

Algorithms Application:- The application of the three algorithms is done on our data set. We tested the algorithms on the same data set using different sizes (600, 800, 1000, 1200) in order to detect the algorithm gives more accuracy and at the same time it guarantees stability. Fig. 7, Fig. 8, Fig. 9 shows the graphical results of the Neural Networks algorithm, Svm Algorithm and KNN Algorithm respectively.



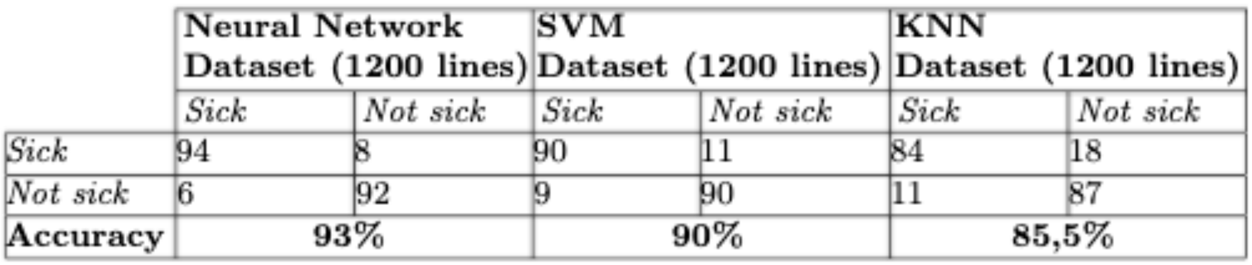
SVM application



KNN Application

Testing algorithms

After running the 3 algorithms on the 4 data sets (600, 800, 1000 and 1200 lines), we test the accuracy of each algorithm on the different data sets. Table 3 illustrates how we calculated the accuracy from the confusion matrix (we took the case of the last data set which contains 1200 lines), on the 3 algorithms. Now we can display the accuracy of each algorithm on the 4 data sets, in

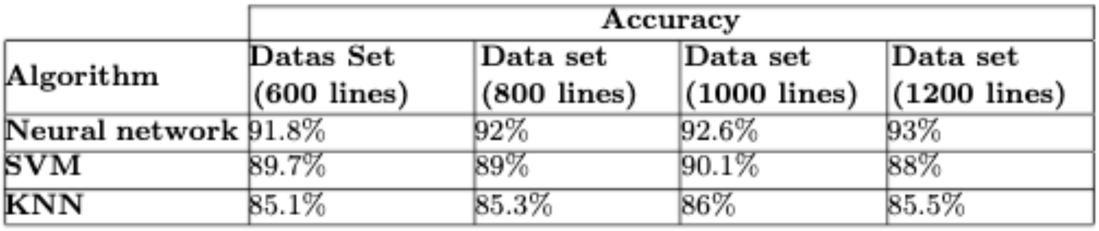


	Neural Network Dataset (1200 lines)		**SVM Dataset (1200 lines)**		**KNN Dataset (1200 lines)**	
	Sick	*Not sick*	*Sick*	*Not sick*	*Sick*	*Not sick*
Sick	94	8	90	11	84	18
Not sick	6	92	9	90	11	87
Accuracy	**93%**		**90%**		**85,5%**	

shows a comparison between the 3 algorithms in terms of accuracy and stability against changes in the data sets.

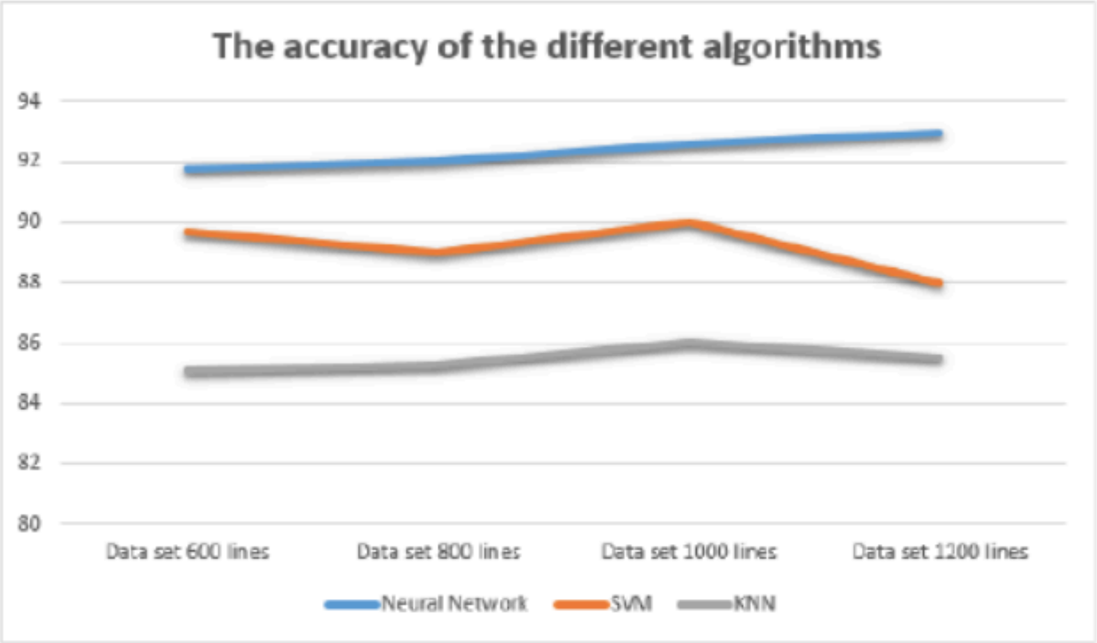
Choosing the best algorithm

After analyzing the results found previously, we find that neural networks is the best algorithm in our study, it is always stable in its results, and gives the best accuracy.



	Accuracy			
Algorithm	Datas Set (600 lines)	Data set (800 lines)	Data set (1000 lines)	Data set (1200 lines)
Neural network	91.8%	92%	92.6%	93%
SVM	89.7%	89%	90.1%	88%
KNN	85.1%	85.3%	86%	85.5%

Comparing the accuracy of different algorithms



9. RESULT AND DISCUSSION

Efficiency of all the machine learning algorithms is shown. Different models based on those efficiency are designed to predict the cardiac disease in a person.

We have trained it (find the patterns) on the training set and tested it (use the patterns) on the test set.

We have tried different machine learning models:

1. Logistic Regression
2. K-Nearest Neighbours Classifier
3. Support Vector machine
4. Decision Tree Classifier
5. Random Forest Classifier

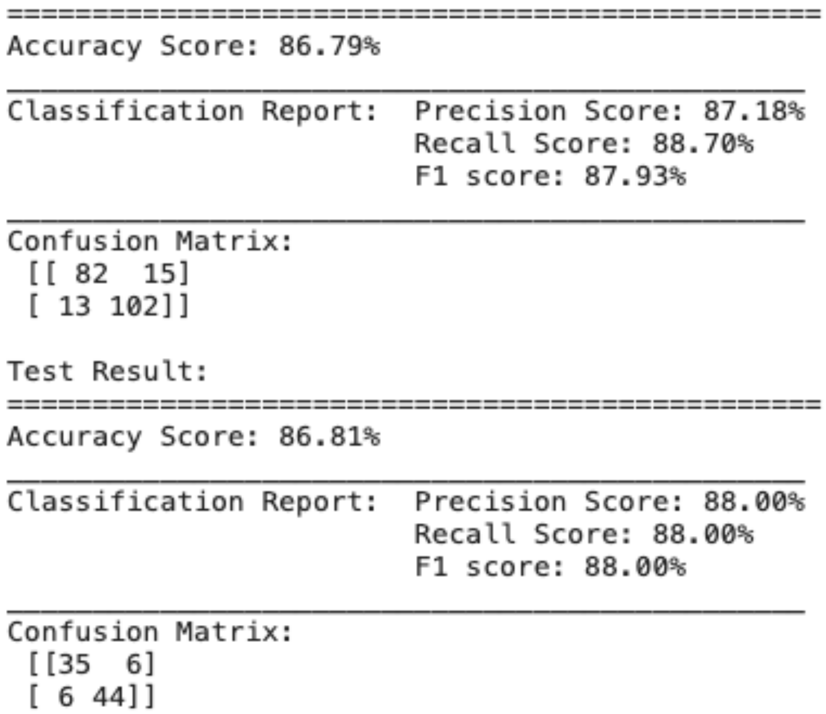
• K-nearest neighbors

```
from sklearn.neighbors import KNeighborsClassifier

knn_classifier = KNeighborsClassifier()
knn_classifier.fit(X_train, y_train)

print_score(knn_classifier, X_train, y_train, X_test, y_test, train=True)
print_score(knn_classifier, X_train, y_train, X_test, y_test, train=False)
```

Train Result



```
================================================
Accuracy Score: 86.79%
_______________________________________________
Classification Report:  Precision Score: 87.18%
                        Recall Score: 88.70%
                        F1 score: 87.93%
_______________________________________________
Confusion Matrix:
 [[ 82  15]
 [ 13 102]]

Test Result:
================================================
Accuracy Score: 86.81%
_______________________________________________
Classification Report:  Precision Score: 88.00%
                        Recall Score: 88.00%
                        F1 score: 88.00%
_______________________________________________
Confusion Matrix:
 [[35  6]
 [ 6 44]]
```

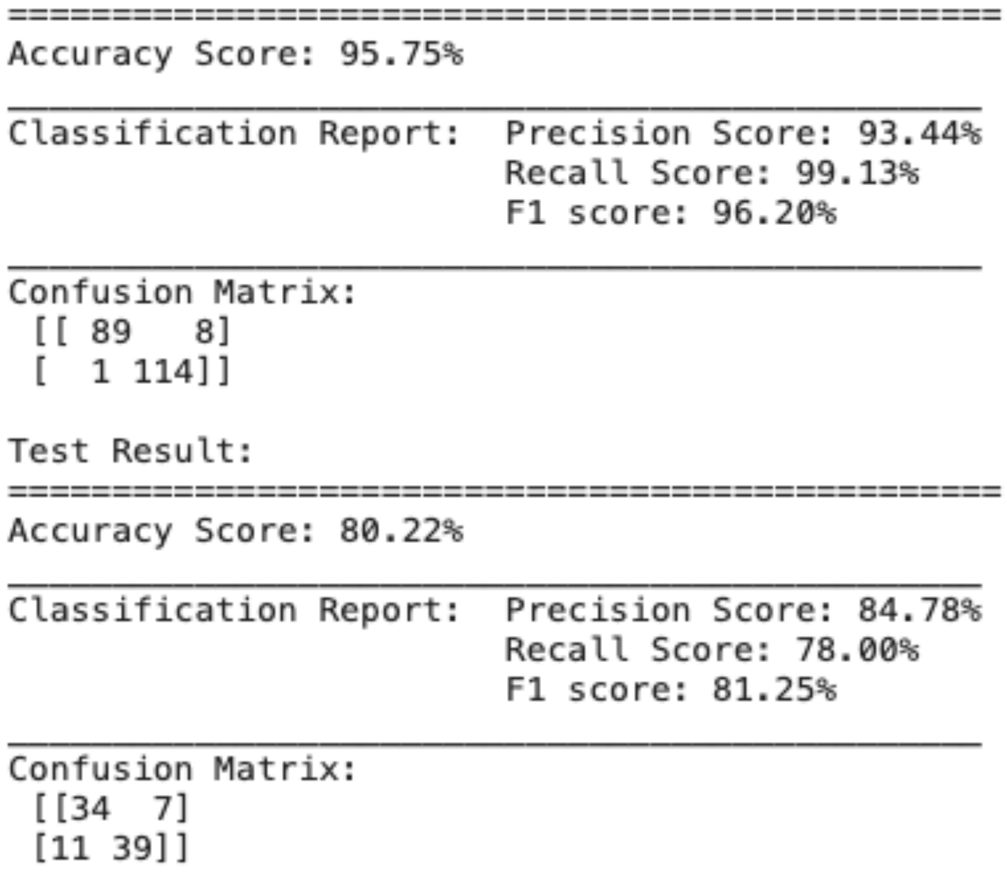
- **Decision Tree Classifiers**

```
from sklearn.tree import DecisionTreeClassifier

tree = DecisionTreeClassifier(criterion = "entropy",max_depth = 5,random_state=1)
tree.fit(X_train, y_train)

print_score(tree, X_train, y_train, X_test, y_test, train=True)
print_score(tree, X_train, y_train, X_test, y_test, train=False)
```

Train Result



```
================================================
Accuracy Score: 95.75%
_______________________________________________
Classification Report:  Precision Score: 93.44%
                        Recall Score: 99.13%
                        F1 score: 96.20%
_______________________________________________
Confusion Matrix:
 [[ 89   8]
 [  1 114]]

Test Result:
================================================
Accuracy Score: 80.22%
_______________________________________________
Classification Report:  Precision Score: 84.78%
                        Recall Score: 78.00%
                        F1 score: 81.25%
_______________________________________________
Confusion Matrix:
 [[34  7]
 [11 39]]
```

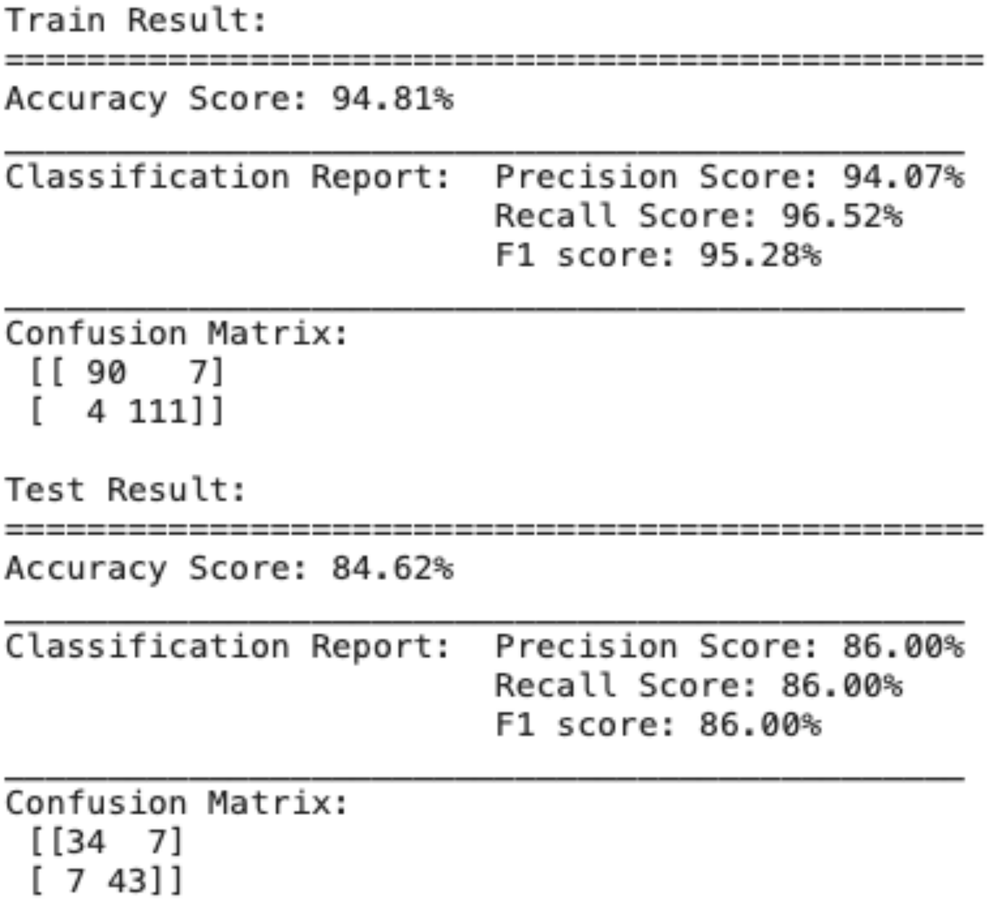
- **Random Forest Classifier**

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import RandomizedSearchCV

rand_forest = RandomForestClassifier(n_estimators=100,max_depth=5, random_state=42)
rand_forest.fit(X_train, y_train)

print_score(rand_forest, X_train, y_train, X_test, y_test, train=True)
print_score(rand_forest, X_train, y_train, X_test, y_test, train=False)
```

Train Result



```
Train Result:
================================================
Accuracy Score: 94.81%
_______________________________________________
Classification Report:  Precision Score: 94.07%
                        Recall Score: 96.52%
                        F1 score: 95.28%
_______________________________________________
Confusion Matrix:
 [[ 90   7]
 [  4 111]]

Test Result:
================================================
Accuracy Score: 84.62%
_______________________________________________
Classification Report:  Precision Score: 86.00%
                        Recall Score: 86.00%
                        F1 score: 86.00%
_______________________________________________
Confusion Matrix:
 [[34  7]
 [ 7 43]]
```

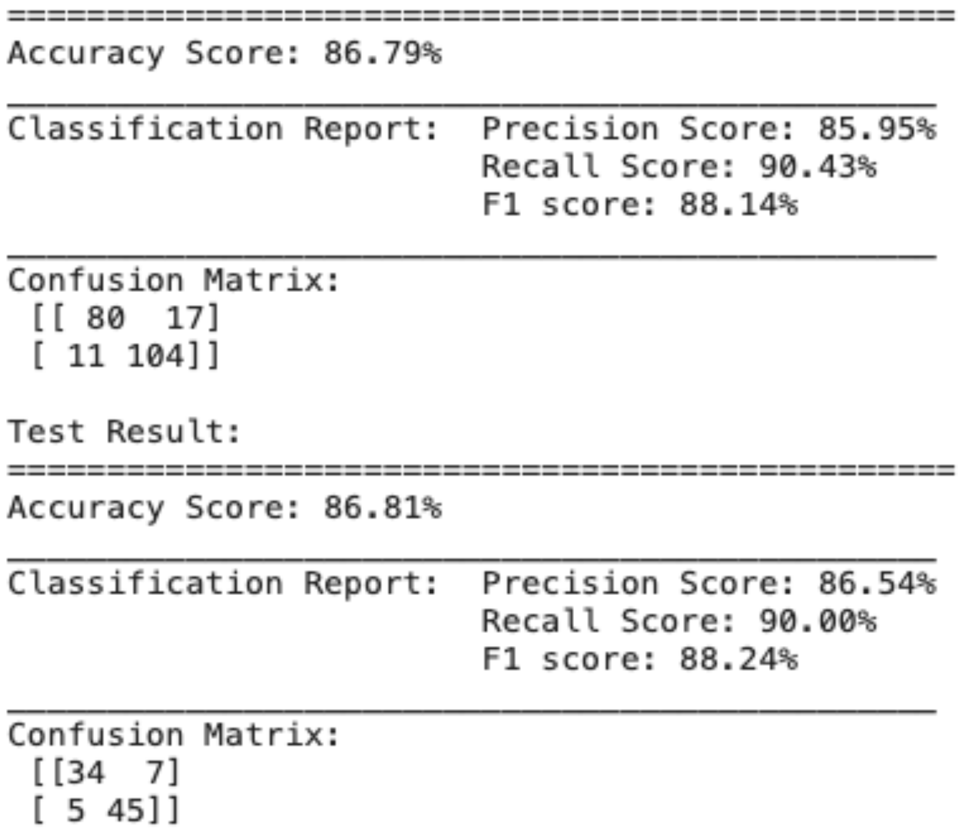
- **Logistic Regression**

```
from sklearn.linear_model import LogisticRegression

log_reg = LogisticRegression(solver='liblinear')
log_reg.fit(X_train, y_train)

print_score(log_reg, X_train, y_train, X_test, y_test, train=True)
print_score(log_reg, X_train, y_train, X_test, y_test, train=False)
```

Train Result



```
================================================
Accuracy Score: 86.79%
_______________________________________________
Classification Report:	Precision Score: 85.95%
			Recall Score: 90.43%
			F1 score: 88.14%
_______________________________________________
Confusion Matrix:
 [[ 80  17]
 [ 11 104]]

Test Result:
================================================
Accuracy Score: 86.81%
_______________________________________________
Classification Report:	Precision Score: 86.54%
			Recall Score: 90.00%
			F1 score: 88.24%
_______________________________________________
Confusion Matrix:
 [[34  7]
 [ 5 45]]
```

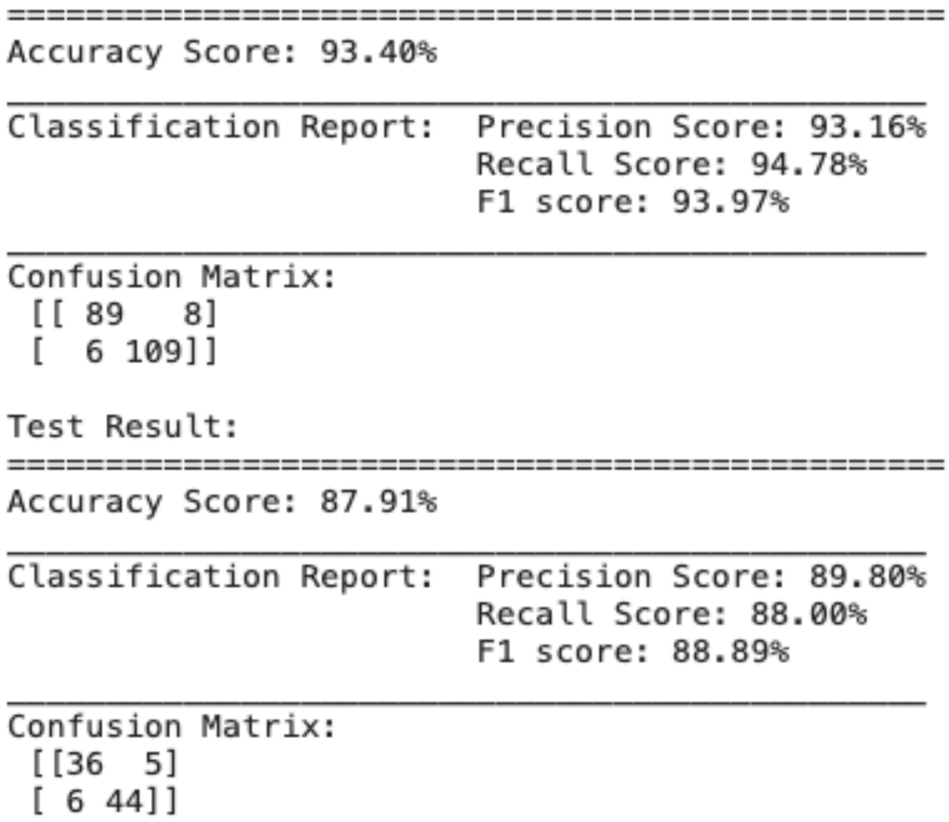
- **Support Vector Machine**

```
from sklearn.svm import SVC

svm_model = SVC(kernel='rbf', gamma=0.1, C=1.0)
svm_model.fit(X_train, y_train)

print_score(svm_model, X_train, y_train, X_test, y_test, train=True)
print_score(svm_model, X_train, y_train, X_test, y_test, train=False)
```

Train Model



```
================================================
Accuracy Score: 93.40%
_______________________________________________
Classification Report:  Precision Score: 93.16%
                        Recall Score: 94.78%
                        F1 score: 93.97%
_______________________________________________
Confusion Matrix:
 [[ 89   8]
 [  6 109]]

Test Result:
================================================
Accuracy Score: 87.91%
_______________________________________________
Classification Report:  Precision Score: 89.80%
                        Recall Score: 88.00%
                        F1 score: 88.89%
_______________________________________________
Confusion Matrix:
 [[36  5]
 [ 6 44]]
```

Accuracy Table of models

The prediction accuracy of different models is given:

Model	**Training Accuracy %**	**Testing Accuracy %**
K-nearest neighbors	86.79%	86.81%
Decision Tree Classifier	95.75%	80.22%
Random Forest Classifier	94.81	84.62%
Logistic Regression	86.79%	86.81
Support Vector Machine	93.4	87.91

10. CONCLUSION

Heart diseases have become more and more frequent among people including our country. Therefore, predicting the disease before becoming infected decreases the risk of death. This prediction is an area that is widely researched. Our paper is part of the research on the detection and prediction of heart disease. It is based on the application of Machine Learning algorithms, of which we have chosen the 5 most used algorithms (KNN, Decision Tree, Random Forest, Logistic Regression and SVM), on a real data set of some people, where we had very good results This study has been successful in establishing the automatic diagnosis of cardiac septal defects. We developed the proposed system to predict heart disease. Ensemble methods (boosting and bagging) with feature extraction algorithms (PCA and LDA) are used to improve predicting heart disease performance. The feature extraction algorithms are used to extract essential features from the Cleveland heart disease dataset. Comparison between ensemble methods (boosting and bagging) and five classifiers (KNN, SVM, NB, DT, and RF) is applied to selected features. The experimental results showed that the bagging ensemble learning algorithm with DT and PCA feature extraction method had achieved the best performance.

The CNN-based architecture can successfully account for segmentation of cardiac defects.. In this study, it was proven that the proposed architecture model has a very high degree of accuracy with a very small error rate for predicting contour lesions in cardiac defects. The preciseness of KNN, Decision Tree, Random Forest, Logistic Regression and SVM for predicting heart disease are 86.79%, 95.75%, 94.81%, 86.79% and 93.4% respectively. Through these findings, the diagnosis of a Cardiac defect will be more precise, so it can be utilized by all physicians when performing an echocardiogram examination. In the future, this research will be carried out with a greater number of patients and by combining several echocardiogram views.

11. REFERENCE

[1] K. Puri, H.D. Allen, A.M. Qureshi
Congenital heart disease
Pediatr Rev, 38 (10) (2017), pp. 471-486, 10.1542/pir.2017-0032

[2] G. Mcleod, et al.
Echocardiography in congenital heart disease
Prog Cardiovasc Dis (2018), 10.1016/j.pcad.2018.11.004
468–75
Google Scholar

[3] A.D.R. Fulton, S. Saleeb
Isolated ventricular septal defects in infants and children : anatomy , clinical features , and diagnosis
(2020), pp. 1-26
Google Scholar

[4] A.G.W. Vick, L.I. Bezold
Isolated atrial septal defects (ASDs) in children: classification, clinical features, and diagnosis
(2020)
Google Scholar

[5] Q.M. Zhao, C. Niu, F. Liu, L. Wu, X.J. Ma, G.Y. Huang
Accuracy of cardiac auscultation in detection of neonatal congenital heart disease by general paediatricians
Cardiol Young, 29 (5) (May 2019), pp. 679-683,
10.1017/S1047951119000799

[6] T.S.M. Kusunose, K. Haga, A. Abe
Utilization of artificial intelligence in echocardiography
Circ J, 83 (8) (2019), pp. 1623-1629, 10.1253/circj.CJ-19-0420
Google Scholar

[7] E. Smistad, A. Ostvik, B.O. Haugen, L. Lovstakken
2D left ventricle segmentation using deep learning
IEEE International Ultrasonics Symposium, IUS (2018),
10.1109/ULTSYM.2017.8092812
1–4
Google Scholar

[8] A. Madani, R. Arnaout, M. Mofrad, R. Arnaout
Fast and accurate view classification of echocardiograms using deep learning
npj Digital Medicine, 1 (1) (2018), pp. 1-8,
10.1038/s41746-017-0013-1

Google Scholar