Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 12791 - 12813
Received 25 April 2021; Accepted 08 May 2021.

Performance and Comparison of Classification Algorithms of MLwith Comparative Mean for Heart Disease Prediction

[1]NehaVerma, [2] Dr. B.P Singh

[1]Scholar PhD, [2]Professor
[1,2]Department of Computer Science Capital University, Jharkhand

Abstract:
The prediction of neurodegenerative diseases and movement disturbances is focused on soft computing. In comparison to the nucleus, the remainder of the body are higher priority sections. It brings oxygen in the body. Data exploration may be used to estimate the distribution of cardiac diseases among medical professionals. Data collection enables medical centers to assess different diseases and allows them to evaluate growing diseases. The main purpose of this research paper is to compare the efficacy of different classifications for cardiac disease prediction. In - month, several patient descriptions are released. The stored data was used to monitor future epidemics. The most significant influence is to anticipate the disease from the current medical studies. Many approaches in the scientific world are being explored to learn more efficient statistics. The value of different learning approaches for the prediction of heart disease is addressed in this study. This report utilizes three different data sets to evaluate the degree of estimation precision. According to the Kaggle and UCI computer data registry, there are more than 250 data in these databases. This paper analyses the efficiency, precision and F1 importance of each data set using numerous algorithms. This thesis analyzed the 11 most efficient classification algorithms, which are Logistic Regression, KNN, Decision tree, Random Forest, SVM, Gaussian NB, Ada Boost Classifier Gradient Boosting Classifier, Quadratic Discriminant Analysis and MLP Classifier with comparative mean of three data sets. This paper investigates the accuracy of prediction from the most concerning algorithms of ML with recall and f-score of the heart data and presented it in table with visual representation.
Keywords: K-Neighbor Classifier, Vector Help Classifier, Random Forest Classifier, Cardiac Attack Predictor, Decision Tree Classifier. Machine Learning (ML)

I. Introduction

In the life of the human race the heart plays a significant function. It provides oxygen to any component of the body. The brain and other tissues will stop to function and fail if the body fails to remain alive for a few minutes. These health issues, lifestyle shifts, workplace stresses and poor eating habits have led to cardiac insufficiency. Heart attack is one of the world's main causes of death. Heart-related diseases can be included in accurate estimates. Different medical institutions gather detailed medical details all over the world. It may use these details to achieve useful views across a variety of soft techniques. However, the data produced is quite large and can be noisy in some circumstances. Its libraries are confusing to navigate utilizing different artificial intelligence approaches. This can be very effective in detecting the occurrence of cardiac conditions.[1]

1.1 Multi-Level Controlled Classification.

Dimensionality implies the preference of mathematical representation, such that the specifics are very important and the trivial facts are left out. A task and problem can need multiple qualities,

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 12791 - 12813
Received 25 April 2021; Accepted 08 May 2021.

but not all of them affect efficiency. There are a number of main variables that may impact program performance, which results in low quality. For a machine learning model, a dimensionality reduction mechanism is needed. Aspect Extraction and Aspect Selection are the main tool for dimension reduction.

A. Function extraction.

In this scenario, a subset from the original subset is extracted. It involves conversion of feature. There is a continuing transfer. This shift ensures that the procedure is not redundant. In order to extract characteristics, primary component analysis (PCA) was used. The key factor analysis is a tool used for statistical analysis. Collect the paths in the feature space with the highest deviation.

B. Function value ranking.

A smaller subset of the original data is picked. CFS works by a mixture of evaluation and search to minimize the dimension. The chi-square test is used to choose the most important functions.[2]

1.2 Coronary Heart Attack.

Soft estimation approaches include strategies for dimensional reduction to reduce the volume of data measurement. This is a key strategy in the analysis of disease-related data sets. This prediction consists primarily of several component sections, from which the data is eventually prediction modelled.

- Shorting of the most relevant details
- Find the data pattern analysis through the ROC curve
- Treatments of lost importance (Replace Mean or median values in vacant spaces)
- Set the data into two sections
- One as a test and the other as a compilation of train data (Prefer 70:30 ration for train: test)
- Apply logistic regression over the sets of data
- Find the exactness of precision
- Find the right algorithm for accuracy.[3-4]

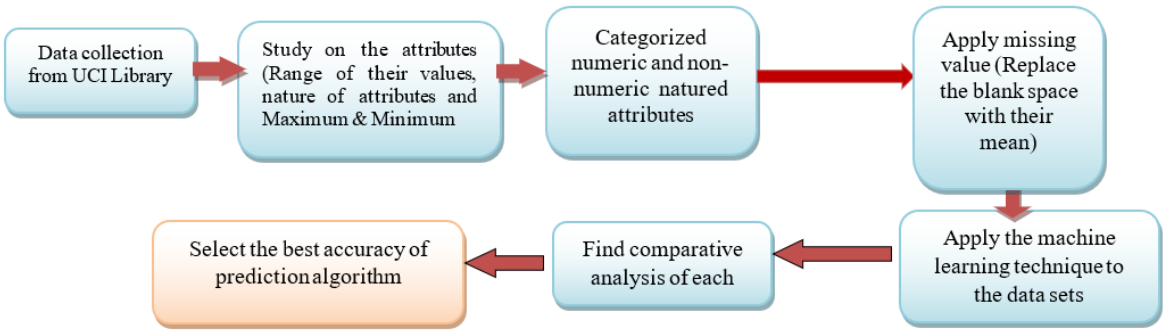


Fig 1. Architecture of the Experiments

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 12791 - 12813
Received 25 April 2021; Accepted 08 May 2021.

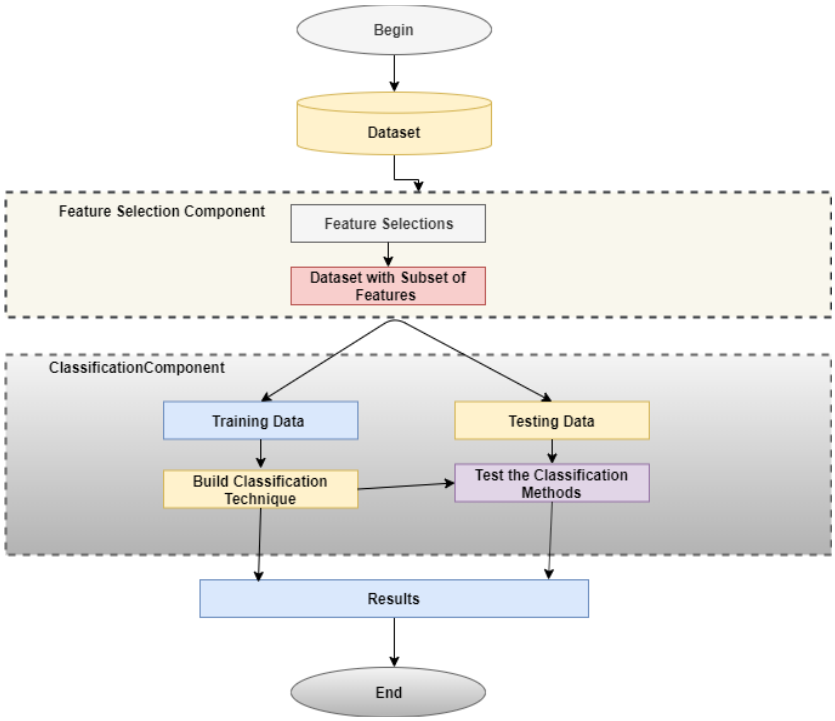


Fig 2: General Flow for Data prediction

II. Background

[1] Long et al. (2015) In this paper they suggest a cardiac diagnostics method based on loosely set reduction of attributes and type 2 interval fuzzy logic framework (IT2FLS). Rough attribute reduction dependent sets and IT2FLS integration are structured to adapt to the complexities and uncertainty of high-dimensional data sets. IT2FLS used a hybrid learning method that involves the modification of parameters with c-fuzzy meaning clustering and unpredictable fire and genetic hybrid algorithms. This is a computer-cost learning method, particularly when used for high-dimensional data sets. Consider the usage of chaotic firefly algorithms to maximize the reduction of roughly specified attributes. This decreases computer sophistication and increases the efficiency of IT2FLS. Experimental findings suggest that the device has major advantages over other Naïve Bayers machine learning approaches, vector support and artificial neural networks. The proposed model should also be used as a decision support method for heart disease diagnosis.

[2] Santhanam&Ephzibah (2015) Health mistakes are commonly costly and dangerous. Every year, they cause many deaths worldwide. Health decision-making processes provide potential for medical mistakes to be minimized and patient care enhanced. The detection and prevention of heart attack is one of the most significant facets of the implementation of this method. Implement classification methods for data mining to examine multiple problems in the heart. This paper aims at establishing a prediction scheme of cardiac disease utilizing the clustering approach of data mining. A health care system is therefore a system rich with numbers. These medical details extract information to forecast illness better. Data mining technology is generally commonly used for the prediction of various diseases in clinical expert systems. These innovations also uncovered associations and trends concealed in medical records. It is also an essential challenge to attempt to help the diagnosis method with the expertise and experience of multi-experts and

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 12791 - 12813
Received 25 April 2021; Accepted 08 May 2021.

the clinical screening evidence of patients in the database. Unfortunately, vast volumes of heart attack data are gathered by the healthcare sector and cannot be accurately diagnosed to detect secret details.

[3]Javed et al. (2018) purpose of this work is to detect heart failure by utilizing machine tools including genetic algorithms and fluid logic. This device allows physicians to simplify cardiovascular disease and patient treatment. They design a hybrid gene that identifies heart failure. For random searches, genetic algorithms are used to have the perfect solution for the practical selection dilemma. The related data set features assist in the creation of a classification model with a fuzzy inference method through the diagnostic system. Sample data produce fuzzy machine laws. Using genetic algorithms in the rule collection to pick important and specific subsets of laws. The proposed research utilizes genomic algorithms and fluid thinking mechanisms to accurately predict cardiac failure in patients. Selected attributes involve sex, serum (chol), maximal heart rate (thalakh), exercise-related angina (exanguine), ST suppression (oldpeak) induced by exercise, main blood vessel amount (ca) or thal. By using the Fuzzy Gaussian membership feature and by using the Centroid approach, machine efficiency can be increased. To aid to explain job quality, work was measured using success metrics such as precision, specificity, sensitivity and uncertainty matrix. The rating accuracy of the layered k-fold system was 86%, with precision and sensitivity values of 0.90 and 0.80. The number of attributes accessible in the UCI Machine learning library in the cardiac disorder dataset has been decreased from 13 to 7. The accuracy of the proposed work is 1.54% higher than the current method. The proposed model is named the GAFL model, a fuzzy logic model for efficient prediction of heart disease. Modeling is easy and provides doctors in clinics and surgical services with a convenient alternative.

[4]Jabbar et al. (2016) Cardiac risk prediction is a huge problem with the large workload, and the prediction of persons with cardiac disease has been the most worried. It is a major struggle to detect the disorder. The concern is that the data are derived with meaningful information. Data mining methods are therefore used to collect useful knowledge. In order to forecast heart attack, decision tree and ID3 are used. Most experts and physicians know about cardiac attack prediction and may use a number of methods to forecast disease. A decision tree is used to forecast heart attack to address this dilemma. This research preprocessed the collected data and a decision tree algorithm and ID3 were used to forecast cardiovascular disease.

[5]Saxena & Sharma (2015),Heart disorder is the world's main cause of early death. It is a difficult challenge to foresee the effect of a disease. Data mining dynamically introduces diagnostic guidelines and lets specialists boost the diagnostic method efficiency. Researchers utilize a number of data retrieval methods to help health staff forecast cardiovascular disease. Random Forest is an integrated and effective medical learning algorithm. Chi-square metrics of selection features are used to test and assess relations between variables. A classification model that uses random forests as a classifier, chi-square approach and genetic algorithm to forecast heart disease is proposed in this paper. The experimental findings revealed that their approach increases classification specificity relative to other classification approaches and that medical practitioners would effectively use the proposed model for forecasting cardiac disease.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 12791 - 12813
Received 25 April 2021; Accepted 08 May 2021.

[6]Sharmila & Gandhi (2017),A big source of morbidity and death is cardiovascular disease (CVD). Identification of cardiovascular disorders is important, but it must be achieved with considerable care and reliability, and it is a difficult challenge to ensure proper automation. Nobody should get the same credentials as a doctor. Both doctors cannot have the same qualifications in all sub-professionals and doctors have convenient access to technical skills in certain places. Automated medical diagnostic devices boost medical treatment and minimize prices. In this research, they have established a framework that effectively detects rules centered on health parameters to forecast patient risk. The priority of the law may be calculated by user specifications. The evaluation of device efficiency based on rating accuracy indicates that the method can more reliably estimate the likelihood of heart failure.

[7]Haq et al. (2018)heart disease is today one of the world's major causes of death. Cardiovascular disorder prediction is an important topic for clinical data review. Machine Learning (ML) has shown itself to help make recommendations and projections based on the large volume of data the healthcare sector produces. They have seen the usage of ML technologies in all aspects of the Internet of Things in recent advancements (IoT). Different experiments only include details on the usage of ML technologies for heart disease prediction. This paper suggests a modern method to boost the predictive performance of cardiovascular disorders by utilizing machine learning techniques. A predictive model was implemented with different combinations of characteristics and many established classification techniques. The random forest and linear model (HRFLM) prediction model will boost efficiency with 88: 7% precision.

[8] Abdaret al. (2015)suggest a new detector on the basis of the transforming coefficients achieved via the point propagation function built by orthogonal polynomials of Chebyshev. Rims close to Prewitt and Roberts were found by the edge detector. Responsive to a parameter¨ adjustable, which can be determined by the conversion factor. They use an edge detector to remove portions of the brain from the human skin scanned magnetic resonance imaging (MRI).

[9]Shinde et al. (2017) Heart disorder is one of the most critical human illnesses in the world that has some severe implications for human health. In cardiovascular disease, the heart cannot push the blood needed to other areas of the body. For the prevention and treatment of heart failure, correct and prompt detection of heart disease is critical. Diagnosing conventional patient history of cardiac failure is in many cases deemed inaccurate. Non-invasive approaches (e.g., machine learning) are accurate and efficient in the classification of healthier individuals and cardiac attack patients. The proposed research developed a machine-based predictive cardiovascular diagnostic device with data from heart disease. Seven machine-based learning algorithms were used, three algorithms for feature selection, cross-validation methods and seven metrics for classificatory efficiency, such as precision, species, sensitivity, Matthews' correlation coefficient and runtime. The 0e method suggested enables the recognition and separation of cardiac patients from healthy persons. Furthermore, the recipient positive curve and region under the curve are determined for each classifier. Both classifiers, feature selection algorithms, preprocessing methods, validation methods and classifying assessment measurements used in this paper were listed. The 0e efficiency of the proposed framework is checked with a complete and streamlined feature set. The decrease of the 0e feature affects the classifier's output about classifier accuracy and runtime. The 0e machine-based decision support device offers doctors with an accurate evaluation of cardiac patients.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 12791 - 12813
Received 25 April 2021; Accepted 08 May 2021.

[10]Gandhi & Singh (2015) Data mining methods have been thoroughly investigated throughout the background of medical data, and prediction of cardiac diseases has proved very important in medicine. Medical background statistics have proved to be heterogeneous, and it suggests that multiple types of data are needed to predict the cardiac condition of a patient. Different strategies of data analysis have been applied to forecast heart attack patients. However, data mining approaches do not eradicate data complexity. Ambiguity in the estimation data was attempted to reduce ambiguity. Membership features are structured to minimize ambiguity and paired with measuring techniques. In addition, an effort was made to identify patient's dependent on medical characteristics. The K-NN classifier minimum gap is combined to distinguish data between classes. You see that the K-NN fuzzy classifier is really strong in contrast to other hardware parameter classifiers.

[11]Otoom et al. (2015.) Heart failure in the United States is one of the largest deaths and morbidity rates. Data mining technologies can estimate a patient's risk of heart attack. The purpose of this analysis was to compare the forecasts for different heart disease data mining algorithms. This task applies and contrasts approaches of data mining to estimate the likelihood of heart attack. Following function study, models of five algorithms, namely C5.0, neuronal network, vector support machine (SVM), K-Nearest neighbors (KNN) and logistic regression, were established and validated. With 93.02 per cent precision, the decision tree C5.0 will create the most reliable model. The KNN, SVM and neural networks account for 88.37%, 86.05% and 80.23%. Decision tree findings are simple to clarify and enforce and multiple practitioners will clearly follow the guidelines.

[12] Parthiban and Srivatsa, (2012),Healing centers, therapeutic services, medical societies produce so much knowledge that they are not utilized properly. The medical sector is not "sufficient in data" but "rich in data." The research approaches to identify associations and trends in the medical details was inadequate. The data mining approach is helpful in this situation. Different data mining methods will also be utilized. The purpose of this white paper is to incorporate numerous abstraction techniques of information utilizing data mining techniques in today's prediction of cardiovascular diseases. This paper analyses data mining techniques for medical databases, such as Naive Bayes, Neural networks and tree decision algorithms.

[13] Dalia M. Atallah et.al [24](2019) The prediction process involves three stages: the DPS phase, the FSS phase and the prediction phase (PS). Both techniques are paired with a modern hybrid sorting process, which selects the minimum number of components that obtain the greatest precision. Finally, it uses the closest neighbours to estimate extreme survival in the classification. The suggested method of prediction was evaluated using the new techniques. Experimental studies have shown that the suggested prediction process beats the new techniques as high precision and limited error F-rate are obtained. This method of prediction may also be used with other input results.

[14] Hoill Jung et.al [25]2013 Proposed an approach that supports a typical pattern therapy judgement to chronic patients. The method suggested is a pain-related decision-making mechanism for chronic condition patients utilizing a traditional sequence medicine for the data processing, extraction and data extraction of standard medical data. Through utilizing simple

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 12791 - 12813
Received 25 April 2021; Accepted 08 May 2021.

patient knowledge to make pain-related choices, frequent changes to the common data mining tree may be created. Pain tends to decide about pain by collecting the same patient details from a trend tree, typically centered on the electronic medical report (EMR).

[15] PavleenKaur et.al [26] 2019 used various machine learning approaches and analyzed public cloud data to create a framework, allowing real-time and remote control of built-in IoT networks and linked to cloud computing. The framework will make recommendations based on historical and pre-cloud evidence. The authors proposed a system for the disclosure of knowledge in the database and for the implementation of transparency which hides trends for the making of sound decisions. This essay discusses prediction mechanisms such as coronary disease, breast cancer, asthma, heart, thyroid, dermatology, liver disease and operative data utilizing several feedback attributes relevant to this individual disease. Experimental findings have been obtained by means of machine learning algorithms such as K-NN, Help Machine, MLP and others used in this report.

III. Algorithms Used

A. Logistic Regression

Regression may be described as calculation and interpretation of the correlation between one or more independent and dependent variables. Regression could be split into two categories: linear and logistic. Logistic regression is prevalent by linear regression. The response variables used mainly to measure binary or multi-class dependent variables are discreet and cannot be modelled explicitly by linear regression. This implies that differential variables are constant values. Logistic regression is used mainly to characterize low-dimensional data at non-linear borders. It also demonstrates the disparity in the proportion of dependent variables and offers a degree depending on value for each variable. The basic and general algorithm for resolving classification problems is logistic regression. It is the same fundamental methodology as linear regression, dubbed "logical regression." The word "logistics" derives from the conceptual function used in this system of classification.The analysis of standard logical functions will start with a logistic regression. A logical function is a sigmoid function which takes a true value from 0 to 1. [6]It has been described as

$$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$$

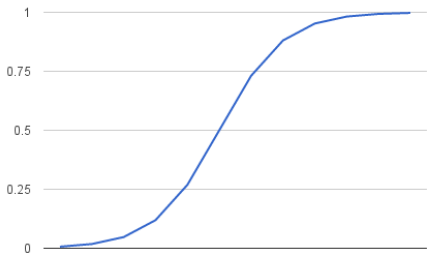


Fig 3: Sigmoid Curve represent the nature of logistic regression

Let's treat t in a univariate regression model as linear function.

$$t = \beta_0 + \beta_1 x$$

This would render the logistic equation

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 12791 - 12813
Received 25 April 2021; Accepted 08 May 2021.

$$p(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

LogisticRegression from sklearn.linear model import
LR = Regression in logistics ()
Fit LR (xtrain.T,ytrain.T)
Print format(LR.score): ("Test Accuracy {}") (xtest.T,ytest.T)
LR = LR.score LR (xtest.T,ytest.T)

B. Support Vector Machine

Supporting vector machines are a very common supervised machine learning technology that can be used as classifiers and predictors utilizing predefined goal variables. Find hyper aircraft which can be categorised in the function space for classification. SVM models represent training data points in the function space and plan them such that they are isolated as far as possible from points from various groups. The test data points are then mapped to the same space and sorted by the side of the margin.[7]

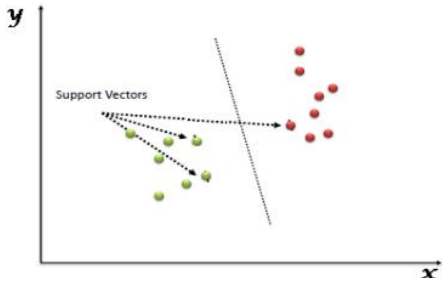


Fig 4: A presentation of SVM

Each input XI has D-Attributes (i.e., dimension D), and is one of two groups Yi = -1 or +1. There are L Training points. In other words, the data format for training is as follows:

$$\{x_i, y_i\}$$

Where$i = 1 \ldots . L, y_i \in \{-1,1\}, x \in \mathcal{R}^D$

This means that the data can be separated linearly. If D = 2, you are able to divide the two groups and draw lines on the x1 and x2 plots that are the x1 map hyperplanes. In D> 2 x2:xD. This hyperplane is describable by w. x + b = 0, where w is perpendicular to hyperplane.b/fundamental to the hyperplane is the vertical gap to the origin. A hyperplane support vector is the closest illustration and the object of the hyperplane support vector machine (SVM) is to put the hyperplane as closely to the members of both groups as possible.[8]

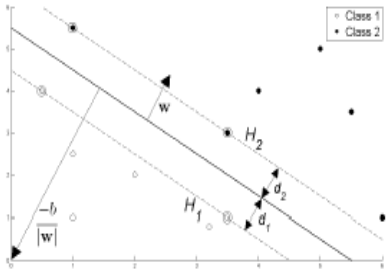


Fig5.: The vector w and b are chosen by two hyperplanes groups to choose w and b to achieve linear separation.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 12791 - 12813
Received 25 April 2021; Accepted 08 May 2021.

$$x_i.w + b \geq +1 \quad for y_i = +1$$
$$x_i.w + b \leq -1 \, for y_i = -1$$

The following can be associated with these equations:

$$yi(x_i.w + b) - 1 \geq 0 \forall_i$$

Taking into account the point closest to the hyperplane, the help vector (in the figure indicated by a circle) may be defined as follows:

$$x_i.w + b = +2 \quad for H_1$$
$$x_i.w + b = -1 \quad for H_2$$

SVC import from sklearn.svm
SVM = SVC(random state=42) SVM
#learning SVM.fit(xtrain.T,ytrain.T)
Test #SVM
Printing ("SVM Accuracy:" (xtest.T,ytest.T)
SVMscore = performance of SVM (xtest.T,ytest.T)
SVM algorithm evaluation accuracy: 86.89 percent.

C. K – Nearest Neighbour

In 1951, Hodges and so on. He implemented a non-parametric model classification system. This is the popular K-Nearest law. One of the simplest powerful grouping methods is K-Nearest Neighbor technology. It is used for classification tasks which do not presume the data usually have little or previous information regarding the dissemination of the data. The algorithm finds the nearest data points in the training set that are similar to the inaccessible data points and an average of the data.[9]

The K-nearest neighbour algorithm essentially reflects plurality voting between the K most close instances of a given "invisible" observation in the classification settings. A similitude is defined on the basis of the distance between two data points. The Euclidean distance is a common option

$$d(x, x') = \sqrt{(x_1 - x'_1)^2 + \ldots + (x_n - x'_n)^2}$$

However, other indicators, such as Manhattan, Chebyshev and Hamming are suitable for particular environments. The KNN classifier carries out the following two steps, provided the positive integer K, the invisible observation x, and similarity d. Calculate d for the whole collection of data between x and each observation of preparation. The K point of training data is considered the nearest point to x of set A. Notice that K is normally quite odd to avoid circumstances of tie.First, determine the conditional likelihood for each group, i.e., the point A scoring for a given category mark. [10-13] (Note I(x) is a feature index. If x is real, the outcome is 1. If not, it's 0.)

$$P(y = j|X = x) = \frac{1}{K} \sum_{i \in A} I(y^{(i)} = j)$$

Finally, our input x is attributed with the greatest likelihood to the class.
= KNNplace(n neighbors = 24) #n neighbors = K value K = n neighbors
#learning model KNNfind.fit(xtrain.T,ytrain.T)
Prediction = NOT find (xtest.T)

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 12791 - 12813
Received 25 April 2021; Accepted 08 May 2021.

Score: {{. score}. Print (25, KNNfind.score) (xtest.T,ytest.T)
KNNscore = KNNfind (xtest.T,ytest.T)

D. Decision Tree

A decision tree is a type of algorithm for supervised learning. This approach is used primarily for issues of grouping. Easily perform for categorical and continuous qualities. The algorithm divides the population into two or more identical sets depending on the main predictors. For each attribute the decision tree algorithm measures the entropy first. The data collection is separated by the highest knowledge benefit or smallest entropy vector or indicator. Ses two measures are achieved recursively for the remaining characteristics.[13-17]

$$Entropy(S) = \sum_{i=1}^{c} -p_i \, log_2 p_i$$

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values\ (A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

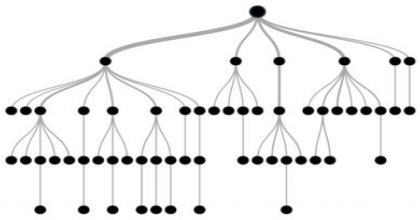


Fig 6: Shows the Decision Tree

Fromsklearn.tree import TreeClassifier Decision
DTC = TreeClassificationDecision(random state=2)
#learning DTC.fit(xtrain.T,ytrain.T)
Pregnancy #
Print ("Decision Tree Score: " (xtest.T,ytest.T)
Score = Score. Score. Score (xtest.T,ytest.T)

E. Random Forest Classifier Methodology

Random Forest is also an algorithm for master learning. This methodology may be used for regression and classification tasks but typically boosts classification tasks efficiency. As the name suggests, before the development, the random forest method considers several decision trees. It is therefore simply a group of decision-making trees. This strategy is based on the conviction that more trees converge on the correct option. For grouping, a vote method is used to evaluate the group, and an average of all outputs from each decision tree is collected for regression. It is ideal for large-scale data sets. [18-21]

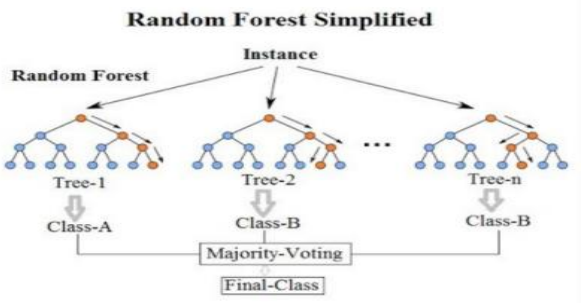


Fig 7: Figure shows the Random Forest

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 12791 - 12813
Received 25 April 2021; Accepted 08 May 2021.

Random Forest (RF) constructs various human decision-making trees through preparation. Summarize all final tree predictions; classification model or average forecast of regression. When making final choices using data, the integration methods are related to.
#n estimator = DT RFfind=RandomForestClassifier(n estimators = 24, random state=5)
RFfind.fit(xtrain.T,ytrain.T)
Print("Random Forest Test Accuracy: " (xtest.T,ytest.T)
Score=Find.scoreRFCscore (xtest.T,ytest.T)
Random Forest Survey Precision: 85.2%

F. Gaussian_NB

The Naive Bayes classification or literally, the Bayesian classification, is based on the theorem of Bayes. The Bayesian network is a special case and a probability-based classifier. Both functions are conditionally autonomous in the Naive Bayes network. Therefore, the improvements in one feature do not impact another feature. The Naive Bayes algorithm can be used to define data sets in large dimensions. The algorithm of the classification uses conditional freedom. Beding isolation means that an attribute value is separate from the meanings of the other class attributes.Let D be a compilation of training data and class labels. Any tuple in the dataset is described by n attributes represented by X={A1, A2, ..., An). Let there be m groups of C1, C2, Cm. For a given tuple X, the classification scheme predicts that X is the class with the greatest posterior likelihood, conditioned by X. [22]Theclassificatory Naive Bayes predicts that Tuple X is Class Ci if and only if

$$P(C_i|X) > P(C_j|X) \quad for \quad 1 \le j \le m, \; j \ne i$$

P (Ci|X) is then maximized. The class Ci for which P is maximized (Ci | X) is considered the post-hypothesis limit. According to the theorem of Bayes,

$$P(C_i|X) \models \frac{P(X|C_i)P(C_i)}{P(X)}$$

If the values of the attribute are conditionally distinct,

$$P(X|C_i) = \prod_{k=1}^{n} P(x_k|C_i)$$

Where xk refers to Ak's value for tuple X. When Ak is categorical, then P (xk |Ci) is the tuple of class Ci in D with xk for Ak divided by |Ci,D|, the number of Ci class tuples in D. The classifier forecasts the class mark of X to be class Ci only if,

$$P(X|C_i)P(C_i) > P(X|C_j)P(C_j) \qquad for \qquad 1 \le j \le m, \; j \ne i$$

Bayesian classifiers are successful in that they have the lowest classification error rate.

Bayes Classification

(Class) = P(Class|Data) * P(Class) (Data)
P(Data)= Trust before
= Gaussian because of normal distribution P(Class|Data)
P(Data) = In NB do not compute this.

Gaussian Naive Bayes

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 12791 - 12813
Received 25 April 2021; Accepted 08 May 2021.

- Ultimately with the Gaussian distribution we have streamlined to eliminate all the squared errors.

- Centered on the law of Bayes, we eventually extracted a square error.

G. Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) (Duda et al., 2001) is a popular methodology used to minimize and classify dimensionality. Provided the number of training images defined by their vectors, we calculate the centroid μi and the covariance matrix for each Class Ci. We presume a standard likelihood of prior class, as in Naïve Bayes. Therefore, we obtain the scatter matrix Sw in class where we attempt to classify each instance into one of 13 groups.[23]

$$Sw = \sum_{i=1}^{13} i$$

LDA allows some simplifying assumptions concerning your data:

- That the knowledge is Gaussian, that each vector is formed like a bell curve.
- If and attribute has the same variation, the values of each element differ by the same average by the same number.
- The LDA model calculates the mean and variance of your data for each class with these assumptions. This is simple to think of in the case of two groups in the univariate (single input variable).

By dividing the sum of values by the total number of values, the mean (mu) value of the input (x) for each class of (k) is usually calculated.

muk = 1/nk * sum(x)

When muk is the mean value of x for class k, nk is the number of class k instances. The variance is measured in both groups as the average quadrated deviation between each value and the mean.

sigma^2 = 1 / (n-K) * sum((x – mu)^2)

Where sigma^2 is the difference between all the inputs (y), n is the number of examples, K is the class number and mu is the mean for input x.

H. Ada_Boost_Classifier

For Adaptive Boosting, AdaBoost is short. Basically, Ada Boosting was the first efficient binary classification boosting algorithm.AdaBoost is a non-linear classification system

- Has strong generalization characteristics: the margin can be proven
- Quite robust to override
- Quite simple to execute

I. Gradient Boosting Classifier

Gradient boosting is one of the competitive algorithms which works on the concept of iteratively boosting weak students by turning their attention to issue observations which have been hard to predict in previous iterations and executing a collection of weak students, generally decision trees. It constructs the model in a step manner, as other boosting approaches do, however generalizes them by optimizing an arbitrary differentiating loss function.[24] Initially we align the model with 75 percent correct observations, and the rest of the unknown variance is reported in the error term:

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 12791 - 12813
Received 25 April 2021; Accepted 08 May 2021.

Y= F(x) + Error

Then we fit another model into the error term in order to add the additional explanatory portion to the initial model, which should increase overall accuracy:

Error = G(x) + Error2

$$F_0(x) = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, \gamma).$$

J. Quadratic Discriminant Analysis

Square discriminant analysis was carried out exactly as in linear discriminant analysis except that we are using the following covariance matrix-based functions for and category:

$Di(X) = -1/2LN(|Si|) -1/2 (X-Y)^{T} Si^{-1} (X-Y)$

Si (X) = di (X) + LN(π)

K. MLP Classifier

An MLP can be regarded as a logistic regression classification in which the data is first processed by a non-linear transformation \Phi. This transition projects the data inserted into a domain in which it can be linearly segregated. This middle layer is considered a secret layer. One secret layer is enough to construct a universal approximator for MLPs. A single secret layer of the MLP (or Artificial Neural Network - ANN) may be graphically represented as follows:

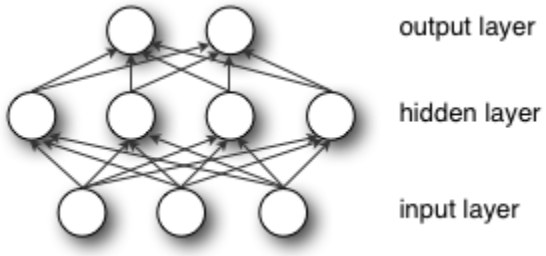


Fig 8: MLP Classifier

Artificial neurons, like secret layers of the multilayer perceptron algorithm, are used in several layers. These algorithms are used for problems of binary classification. For each neuron, a perceptron uses an activation mechanism. Multilayer perceptron's are biological neuron algorithms. You use artificial neurons or perceptron's. The activation function maps each neuron's weighted input and decreases the number of layers to two layers. A perceptron learns from various weights. Below is the algorithm for a multi-layer perceptron.[25-26]

IV. ProposedSteps for Data Modeling

4.1 Procedures

STEP 1: find essential heart data sets attributes. For statistical research, the attribute with minimum and maximum data sets is chosen.

STEP 2: Determine data normality through mathematical review.
STEP 3: Determine the mean and media for the care of lost qualities.
STEP 4: Complete the missed values with media and data set median.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 12791 - 12813
Received 25 April 2021; Accepted 08 May 2021.

STEP 5: Split the test and train research data with a ratio of 70:30.
STEP 6: Execute the train data collection soft learning algorithm
STEP 7: Determine the consistency of the test data sets.

4.2Methodology

Step 1: Dataset Preprocessing
{
Outline of the data
Determine and delete outliers
Identify and process the missing details
Apply effective standardization strategies
Substitute the mean and median
}
}
Step 2: Model collection
{
Data discovery importance (classes)

M-learning learning Sorting algorithm
}
Step 3: Python Model Implementation
{
Import Data Integrating all templates with Python
}
Step 4: Classification Results
{
Accuracy estimation by the operator "Performance" Analyzes outcomes by precise measurement
}
Step 5: Comparison of findings
{
Coping accuracy of all models Comparing the outcome with all M-learning algorithms proposed

Calculate the final performance of each algorithm proposed

Aim for the best in all.
}

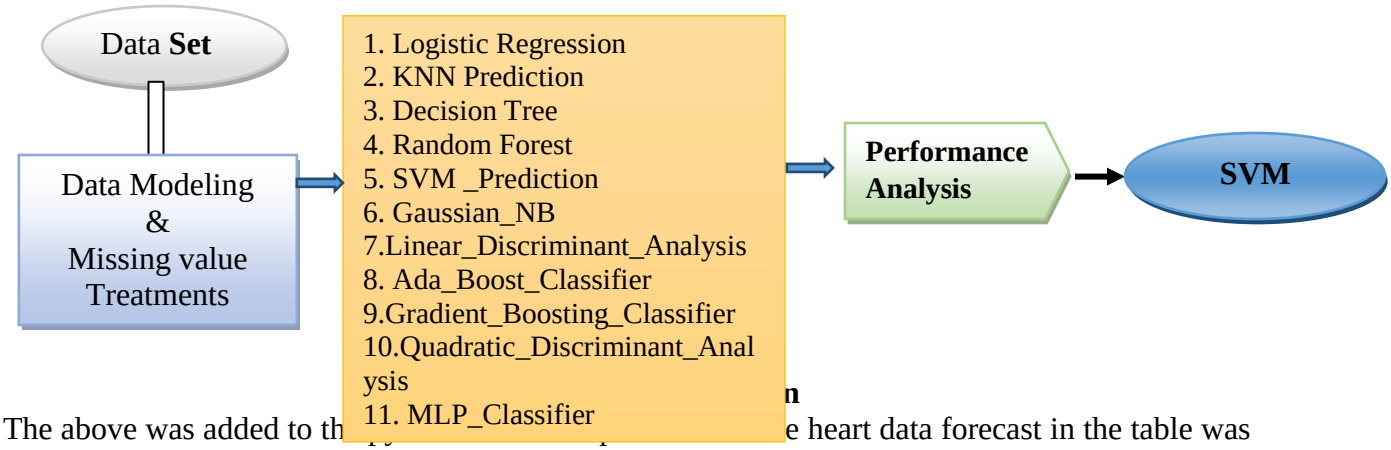
4.3 Pseudo Code

Let a1={a1,q2,a3, … an} be the given dataset
A= {}, the set of Algorithms classifiers
M=Mean and Median {c1, c2, c3, …cn}, the set of
Z =Mean, median of M.
for (i=vacant, i = 0, i++);
{
for (j= vacant, j = 0, j++);

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 12791 - 12813
Received 25 April 2021; Accepted 08 May 2021.

}
Apply M-Learning Algorithm
f = ML (Mod: Data);
Let D={d1,d2,d3, ...dn} be the given dataset
E = {E1, E2, E3, ...En}, the set of ensemble classifiers
C = {c1, c2, c3, ...cn}, the set of classifiers
X = the training set, X D
Y = the test set, Y D
K = meta level classifier
L = n(D)
for i =1 to L do
M(i) = Model trained using E(i) on X
Next i
M=M K
Result = Y classified by M

4.4 Flow Chart of Execution



n

The above was added to th e heart data forecast in the table was determined as follows. This indicates that the SVM algorithm is as reliable as other algorithms.

Exploratory Analysis

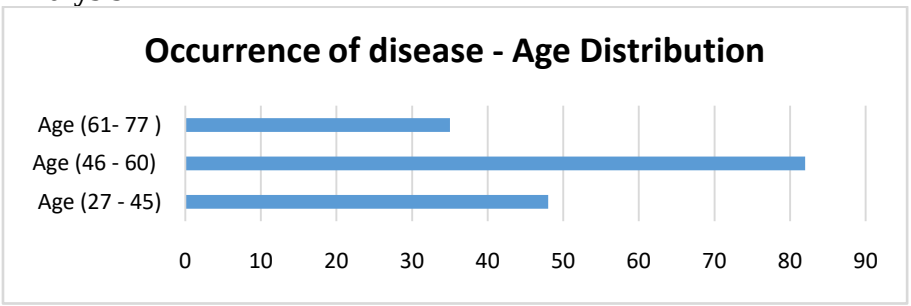


Fig 9: Age wise occurrence of heart disease

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 12791 - 12813
Received 25 April 2021; Accepted 08 May 2021.

The above figure presented the age wise occurrence of heart disease. This analysis is simple distribution of data.

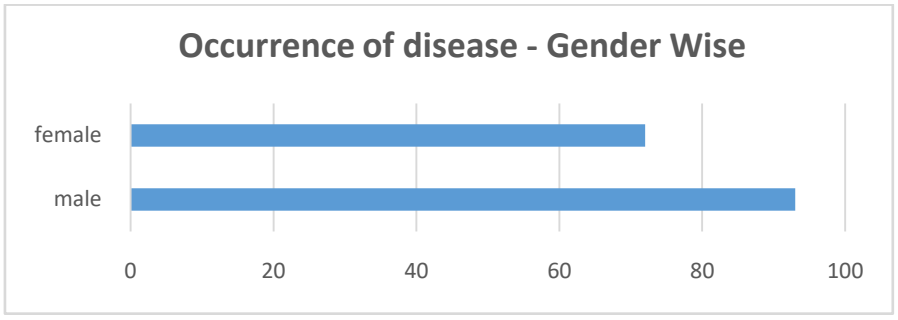


Fig 10: Occurrence of disease - Gender Wise

The above figure presented the Occurrence of disease gender wise. This analysis is simple distribution of data.

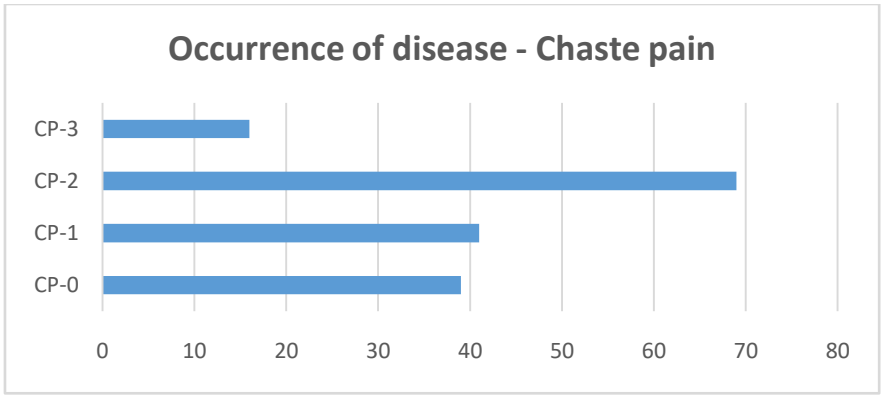


Fig 11: Occurrence of disease - Chaste pain

The above figure presented the Occurrence of disease chaste pain**.** This analysis is simple distribution of data.

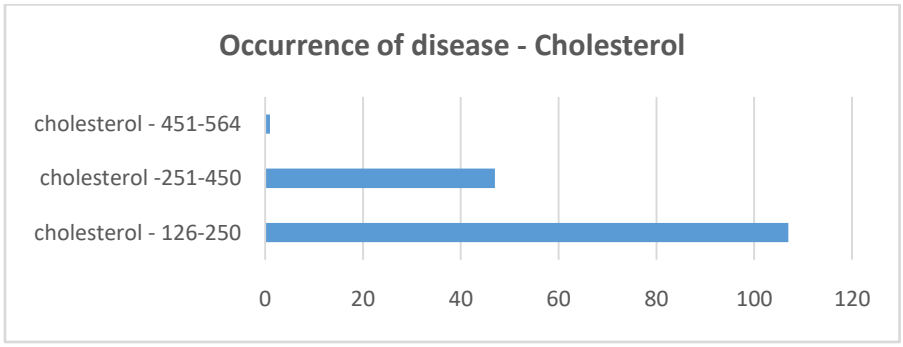


Fig 12: Occurrence of disease - Cholesterol

The above figure presented the Occurrence of disease cholesterol. This analysis is simple distribution of data.

Prediction Analysis

This exploration utilizes various methods to examine the efficiency, accuracy, and F1 significance of each data set. The focus of this study was to conduct an analysis of the 11 most efficient classification algorithms. These algorithms were Logistic Regression, KNN, Decision tree, Random Forest, SVM, Gaussian NB, Ada Boost Classifier Gradient Boosting Classifier,

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 12791 - 12813
Received 25 April 2021; Accepted 08 May 2021.

Quadratic Discriminant Analysis, and MLP Classifier. The outcome of accuracy, Recall and F-measure has been presented in numeric and graphical form.

Classification Rate/ Accuracy:

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

The classification rate or accuracy shall be calculated by the relationship. However, consistency issues remain. On all forms of failures, it means equivalent costs. 99% accuracy will be depending on the issue be excellent, decent, average, bad or awful.

Recall

Recall can be described as the ratio of the total number of positive examples correctly classified by the overall number of good examples. High Recall suggests that the class is remembered correctly (small number of FN).

$$Recall = \frac{TP}{TP + FN}$$

F-measure

Because we have two variables (precision and recall), it helps to compare the two. We evaluate an F-measure that uses Harmonic Mean instead of Arithmetic Mean since it further punishes extremes. The F-message is often closer to the smaller exact or recalled value.[21-26]

Table 1: Comparative table of M- learning algorithms (Data Set 1)

Algorithms	1	2	3	4	5	6	7	8	9	10	11
Accuracy	85.24	73.77	83.61	85.2	**86.89**	**86.89**	**86.89**	80.33	78.69	83.61	83.61
Precision	0.8965	0.9	0.9259	0.8484	0.9	0.9	0.875	0.85	0.806	0.806	0.806
f1-score	0.8524	0.6923	0.8474	0.8615	0.87	0.87	0.875	0.79	0.793	0.793	0.793

1. Regression of the logistics, 2. KNN Prophecy, 3.Tree Judgment, 4. Forest Random, 5. 6. Gaussian NB, 7. Linear Discriminant Analysis, 8.Ada Boost Classification 9. 9. Classifying Boosting Gradient, 10. Quadratic Analysis-Discriminant, 11.MLP Classification.

Data collection 1 has been taken under various classifications for the precision, accuracy, and F1-Score of heart disease prediction. As seen in the above table, all 11 classifiers have been done by machine learning algorithms. The outcome in the table above may easily be evaluated then the accuracy standard for its exact meaning. The result shows that SVM Projection, Gaussian NB and Linear Discriminant Analysis, which is 86.89, are the most accurate supplier. Precision is the most agreed thing that can be thought about for the prediction and the design of the system to choose the right classificatory.

Table 2: Comparative table of M- learning algorithms (Data Set 2)

Algorithms	1	2	3	4	5	6	7	8	9	10	11
Accuracy	77.17	70.65	73.91	72.8	70.65	77.17	75.00	77.17	72.83	76.09	**78.26**

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 12791 - 12813
Received 25 April 2021; Accepted 08 May 2021.

Precision	0.7142	0.6875	0.6744	0.6590	0.6875	0.7045	0.7	0.7045	0.6590	0.6590	0.6590
f1-score	0.7407	0.6197	0.7073	0.6987	0.6197	0.7469	0.7088	0.7469	0.6987	0.6987	0.6987

1. Regression of the logistics, 2. KNN Prophecy, 3.Tree Judgment, 4. Forest Random, 5. SVM _Prompt 6. NB Gaussian, 7. Linear Analysis-Discriminant, 8.Ada Boost Classification 9. 9. Classifying Boosting Gradient, 10. Quadratic Analysis-Discriminant_, 11. MLP Classification

Data collection 2 for the estimation of heart attack dependent on accuracy, precision and F-1 score in multiple classifications. Both 11 classifiers have been done by algorithms of machine learning, as seen in the above table. The findings in the above table display the accuracy of the MLP Classifier, which is 78.26 for the most reliable supplier. Accuracy is the most known problem when selecting the right forecast classifier and setting up a system.

Table 3: Comparative table of M- learning algorithms (Data Set 3)

Algorithms	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Accuracy	**92.59**	81.48	74.07	87.0	81.48	90.74	**92.59**	88.89	75.93	88.89	90.74
Precision	0.9473	1.0	0.64	0.9375	0.7619	0.9444	0.947	0.9411	0.722	0.722	0.7222
f1-score	0.9	0.6875	0.6956	0.8108	0.7619	0.8717	0.9	0.8421	0.666	0.666	0.6666

1. Regression of the logistics, 2. KNN Prophecy, 3.Tree Judgment, 4. Forest Random, 5. SVM _Class 6. NB Gaussian, 7. Linear Analysis-Discriminant, 8.Ada Boost Classification 9. 9. Classifying Boosting Gradient, 10. Quadratic Analysis-Discriminant_, 11. MLP Classification

Data set-3 was graded according to the various precision, accuracy and first grade for heart conditions prediction and all eleven compilations were applied by machine learning algorithms as seen in the above chart. The table can be conveniently evaluated from the precision stage, and the above findings indicate that Logistic Regression and Linear Discriminant Analysis are the top provider for both. The specificity is the best thing to remember when selecting the best grouping for prediction and construction frameworks.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 12791 - 12813
Received 25 April 2021; Accepted 08 May 2021.

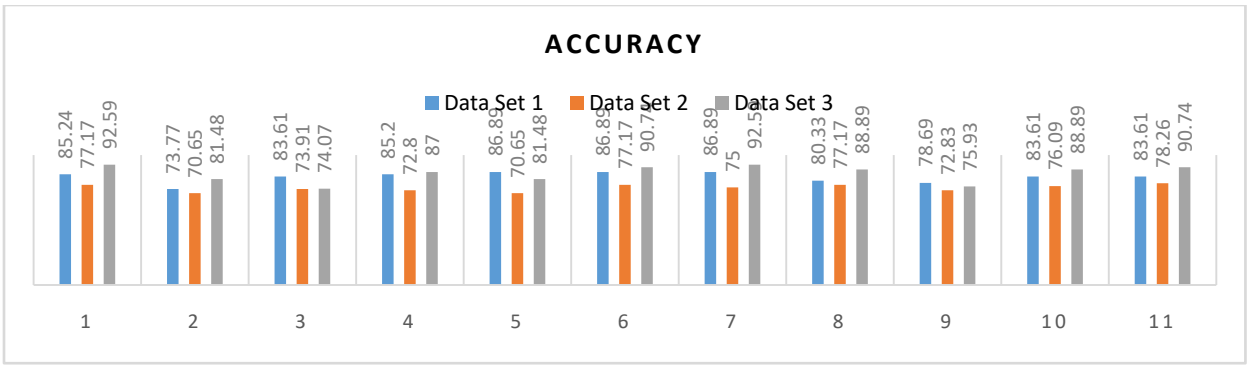


Fig 13: Shows the comparative result of accuracy of prediction.

The graph above demonstrates the comparative accuracy study of three data sets. Linear Discriminant Analysis in the data set 1, SVM Prediction Gaussian NB, has found best accuracy of 86.89, MLP Classification in data set 2 has found best precision 78.26, and Logistic Regression and Linear Discriminant Analysis has found maximum accuracy providers in the data set.

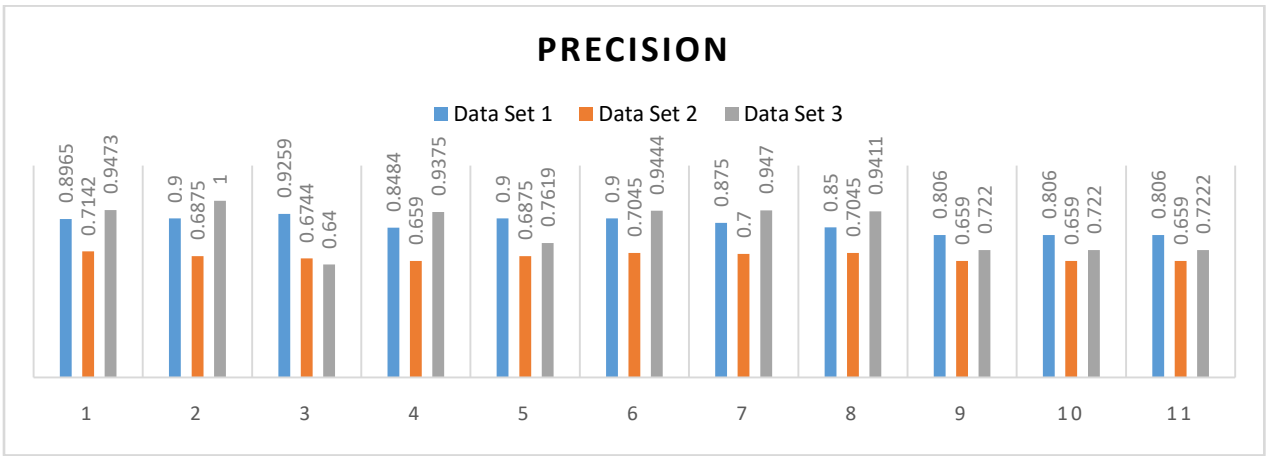


Fig 14: Shows the comparative result of Precision of prediction.

The graph above demonstrates the comparative accuracy study of three data sets. In data set 1, Decision Tree found the maximum precision value 92.59 and in data set 2, logistic regression found the best accuracy value in data set (71.42).

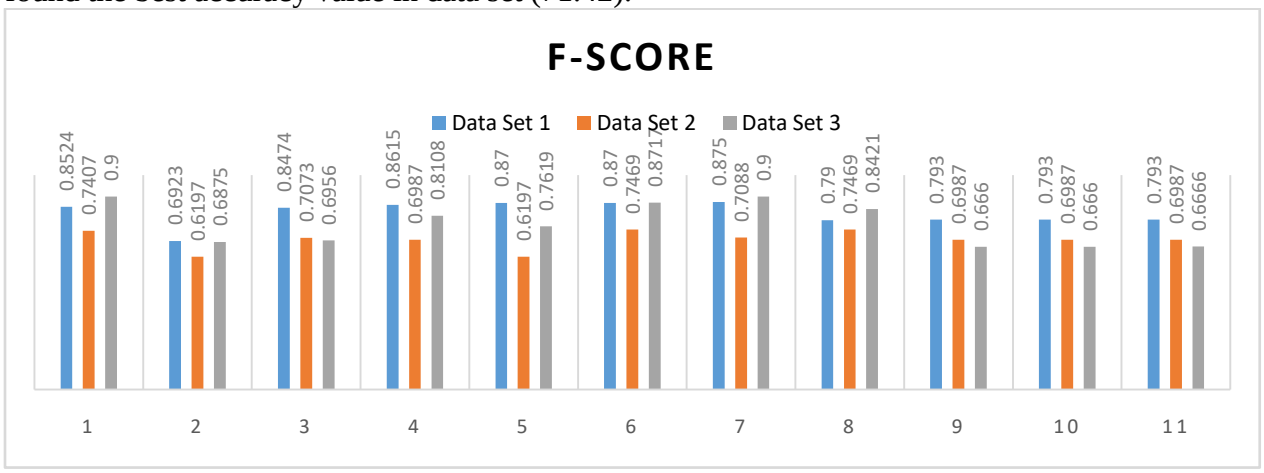


Fig 15: Shows the comparative result of F-Score of prediction.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 12791 - 12813
Received 25 April 2021; Accepted 08 May 2021.

The graph above demonstrates the comparative accuracy study of three data sets. The best F-score value was identified in the data set-1, SVM estimation, the Gaussian-NB and the Linear Discriminant analysis, 87.75, in data set-2, Logistic Regression, Gaussian-NB and Ada Improve classificatory, and the best F-score value was found at 74.07 and the maximum value provider was found in the data set -3 Logistic regression and Linear classification analysis.

Table 4: Comparison with Existing Research

Model	Year	Techniques	Disease	Tool	Accuracy
Otoom et al.	2015	Bayes Net SVM Functional Trees	Heart	Weka	84.5% (All)
Vembandasamy et al.	2015	Naive Bayes	Heart	Weka	86.4 %
Parthiban et al.	2012	Naive Bayes	Heart	Weka	74.1 %
Latha and Jeeva	2019	Majority vote with NB, BN, RF and MP	Heart Disease	Python	85.48%
Tarawneh&Embarak	2019	Naïve Bayes, SVM, KNN, NN, J4.8, RF, and GA	Heart Disease	Python	89.2%,
Sajeev et al.	2019	DL - Multi-Layer Perceptron	Heart Disease	Python	83.4%
Amin et al.	2018	Vote with Naïve Bayes and Logistic Regression	Heart Disease	Python	87.41%
Chauhan et al.	2018	Decision Tree	Heart Disease	Rapid Miner	75.10%
Desai et al.	2019	BPNN	Heart Disease	Python	85.07%
Dwivedi	2016	k-NN	Heart Disease	Python	0.80%
Gokulnath&Shantharajah	2018	SVM	Heart Disease	MATLAB	88.34%
Maji & Arora	2019	Hybrid-DT	Heart Disease	Weka	78.14%
Dalia M. Atallah et.al	2019	Data mining techniques	Kidney	Python	80.77%
Hoill Jung et.al	2013	decision supporting method	Chronic disease	Python	NA
PavleenKaur et.al	2019	Machine learning	Healthcare	Python	80.1%.
Proposed	2020	Machine Learning Comparative Mean	Heart Data sets	Python	92.59 %

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 12791 - 12813
Received 25 April 2021; Accepted 08 May 2021.

10. Conclusion and Future work

In this paper, we explain some successful approaches for forecasting heart disease and test the precision of the classification methodology on the basis of the algorithm chosen. The creation of reliable and computerized classifiers for medical applications is a major problem in the area of Exploratory Analysis and M- learning. We reviewed the three separate data packages of cardiovascular disturbances through Logistical Regression, KNN Prediction, Random Forest Decision Tree, SVM Prediction,Gaussian NB, Ada Boost Classifier, Radiant Boosting Classifier, Quadratic Discriminant Analysis and MLP Classifier. The table above correlates the classification methodology proposed with previous study findings. This thesis explores the optimal learning algorithm for forecasting cardiac failure utilizing different learning methods. This paper utilizes three separate data sets to assess the exactness by the precision rate of the forecast. This paper explores consistency, accuracy and F1 meaning for different learning algorithms in each data collection. In the future it would be a very challenging job for vast societies of the planet to have physicians to a significant number of individuals. In addition to lifestyle improvements, transformation hashas a significant impact on the lifestyle of the metropolitan community. In this case, it is important to provide an integrated device that will allow doctors to predict the disease. This report reveals the success of numerous strategies of machine learning in the study of cardiac disease and has the strongest predictive analysis for each of the three data sets. The best computer algorithms for each data set give the distinguished result in this report. The application of various machine learning suggests different precision of prediction. This different data collection offers different precision, accuracy and F-score on multiple machines learning algorithms.

References

1. Long, N. C., Meesad, P., & Unger, H. (2015). A highly accurate firefly-based algorithm for heart disease prediction. *Expert Systems with Applications*, *42*(21), 8221-8231.
2. Santhanam, T., &Ephzibah, E. P. (2015). Heart disease prediction using hybrid genetic fuzzy model. *Indian Journal of Science and Technology*, *8*(9), 797.
3. Javed, S., Javed, H., Saddique, A., & Rafiq, B. (2018). Human Heart Disease Prediction System Using Data Mining Techniques. *Sir Syed Research Journal of Engineering & Technology*, *8*(II).
4. Jabbar, M. A., Deekshatulu, B. L., & Chandra, P. (2016). Intelligent heart disease prediction system using random forest and evolutionary approach. *Journal of Network and Innovative Computing*, *4*(2016), 175-184.
5. Saxena, K., & Sharma, R. (2015, May). Efficient heart disease prediction system using decision tree. In *International Conference on Computing, Communication & Automation* (pp. 72-77). IEEE.
6. Sharmila, S., & Gandhi, M. I. (2017). Heart Disease Prediction Using Data Mining Techniques-Comparative Study. *Computational Methods, Communication Techniques and Informatics*, 351.
7. Haq, A. U., Li, J. P., Memon, M. H., Nazir, S., & Sun, R. (2018). A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. *Mobile Information Systems*, *2018*.
8. Abdar, M., Kalhori, S. R. N., Sutikno, T., Subroto, I. M. I., &Arji, G. (2015). Comparing Performance of Data Mining Algorithms in Prediction Heart Diseases. *International Journal of Electrical & Computer Engineering (2088-8708)*, *5*(6).

9. Shinde, A., Kale, S., Samant, R., Naik, A., &Ghorpade, S. (2017). Heart Disease Prediction System using Multilayered Feed Forward Neural Network and Back Propagation Neural Network. *International Journal of Computer Applications*, *166*(7), 32-36.
10. Gandhi, M., & Singh, S. N. (2015, February). Predictions in heart disease using techniques of data mining. In *2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE)* (pp. 520-525). IEEE.
11. F. Otoom, E. E. Abdallah, Y. Kilani, A. Kefaye, and M. Ashour, (2015.) "Effective diagnosis and monitoring of heart disease", *International Journal of Software Engineering and Its Applications*, Vol.9, No.1, pp. 143-156.
12. G. Parthiban and S. K. Srivatsa, (2012)"Applying machine learning methods in diagnosing heart disease for diabetic patients", *International Journal of Applied Information Systems,* Vol.3, No.7, pp.2249-0868.
13. K. Vembandasamy, R. Sasipriya, and E. Deepa, (2015)"Heart Diseases Detection Using Naive Bayes Algorithm", *IJISET-International Journal of Innovative Science, Engineering & Technology*, Vol.2, pp.441-444.
14. Latha, C. B. C., & Jeeva, S. C. (2019). Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. *Informatics in Medicine Unlocked*, *16*, 100203.
15. Tarawneh, M., &Embarak, O. (2019, February). Hybrid Approach for Heart Disease Prediction Using Data Mining Techniques. In *International Conference on Emerging Internetworking, Data & Web Technologies* (pp. 447-454). Springer, Cham.
16. Sajeev, S., Maeder, A., Champion, S., Beleigoli, A., Ton, C., Kong, X., & Shu, M. (2019). Deep Learning to Improve Heart Disease Risk Prediction. In *Machine Learning and Medical Engineering for Cardiovascular Health and Intravascular Imaging and Computer Assisted Stenting* (pp. 96-103). Springer, Cham.
17. Amin, M. S., Chiam, Y. K., &Varathan, K. D. (2019). Identification of significant features and data mining techniques in predicting heart disease. *Telematics and Informatics*, *36*, 82-93.
18. Burse, K., Kirar, V. P. S., Burse, A., & Burse, R. (2019). Various Preprocessing Methods for Neural Network Based Heart Disease Prediction. In *Smart Innovations in Communication and Computational Sciences* (pp. 55-65). Springer, Singapore.
19. Chauhan, R., Jangade, R., &Rekapally, R. (2018). Classification Model for Prediction of Heart Disease. In *Soft Computing: Theories and Applications* (pp. 707-714). Springer, Singapore.
20. Desai, S. D., Giraddi, S., Narayankar, P., Pudakalakatti, N. R., &Sulegaon, S. (2019). Back-propagation neural network versus logistic regression in heart disease classification. In *Advanced Computing and Communication Technologies* (pp. 133-144). Springer, Singapore.
21. Dwivedi, A. K. (2018). Performance evaluation of different machine learning techniques for prediction of heart disease. *Neural Computing and Applications*, *29*(10), 685-693.
22. Gokulnath, C. B., &Shantharajah, S. P. (2019). An optimized feature selection based on genetic approach and support vector machine for heart disease. *Cluster Computing*, *22*(6), 14777-14787.
23. Maji, S., & Arora, S. (2019). Decision Tree Algorithms for Prediction of Heart Disease. In *Information and Communication Technology for Competitive Strategies* (pp. 447-454). Springer, Singapore.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 12791 - 12813
Received 25 April 2021; Accepted 08 May 2021.

24. Kaur, P., Kumar, R., & Kumar, M. (2019). A healthcare monitoring system using random forest and internet of things (IoT). *Multimedia Tools and Applications*, *78*(14), 19905-19916.
25. Jung, H., Chung, K. Y., & Lee, Y. H. (2015). Decision supporting method for chronic disease patients based on mining frequent pattern tree. *Multimedia Tools and Applications*, *74*(20), 8979-8991.
26. Atallah, D. M., Badawy, M., El-Sayed, A., &Ghoneim, M. A. (2019). Predicting kidney transplantation outcome based on hybrid feature selection and KNN classifier. *Multimedia Tools and Applications*, *78*(14), 20383-20407