MACHINE LEARNING MODELS FOR HEART DISEASE PREDICTION-A REVIEW

Parvati Kanaki

Research Scholar, REVA University, Bengaluru, Karnataka, India
parvatis2829@gmail.com

Dr Gyanappa A. Walikar

Professor, School of Electronics and Communication Engineering,
REVA University, Bengaluru, Karnataka, India.
gyanapp@rediffmail.com

Abstract

Heart disease is one of the most challenging tasks in health care. Several works have been proposed to address this challenge. In the recent past, some of the research articles are focused on the design of different machine learning algorithms for heart disease in health care, in this paper, we present an overview of supervised, unsupervised, and reinforcement learning algorithms used in the prediction of heart disease wherever possible. This review paper classifies some machine learning algorithms into different categories named Naïve Bays, Decision trees, support vectors, KNN, random forest, genetic algorithms, and structures with their relative performances. This paper also compares different algorithm techniques for heart diseases against accuracy and reliability.

Keywords: **Heart Disease Prediction, ML Algorithms, Logistic Regression, Random Forest, SVM, Genetic algorithm.**

1. Introduction

Researchers are quite worried about heart disease, among which one of the biggest issues is accurately detecting and locating its existence within a person. Even medical professors are not very good at predicting cardiac illness; therefore, early procedures haven't been very effective in discovering it [1]. For forecasting cardiac disease, a variety of medical devices are on the market. They have two main issues: first, they are exceedingly costly, and second, tests are ineffective at estimating the likelihood that a person will develop heart disease. There is a huge need for study in the field of forecasting cardiovascular disease in humans since, in accordance with the most recent WHO questionnaire, just six percent of heart illnesses could be successfully anticipated by medical professionals [2]. Medical technology is one of the sectors where the application of computer science may be applied since growth in the domain has opened up a wide range of prospects. The range of computer science applications ranges from measurement to ocean engineering. The field of medicine has used many of the most significant computer science techniques; during the last ten years, AI has come into its own. ML is one method that is often used across a variety of industries since it does not need a new algorithm for each collection. ML's adjustable capabilities are quite powerful and provide many new opportunities for fields like healthcare.

Heart disease remains one of the greatest issues in medical research since it involves many factors and sophistication to effectively predict this condition. Since this versatile tool uses vector features and different kinds of data under different circumstances for establishing the possibility of cardiovascular disease algorithms like NB, DT, KNN, and Neural Networks are utilized to expect the possibility of peripheral arterial/coronary heart ailments. However, ML may be a superior option to achieve excellent precision for forecasting not merely cardiovascular disease, and other medical conditions. Every algorithm has a unique area of expertise, like how Naive Bayes uses probabilities to predict heart illness, Decision Trees give categorized reports for heart disease, and Neural Networks offer possibilities to reduce heart disease prediction inaccuracy.

All these methods use historical patient records to predict future outcomes for new patients. Many lives may be saved because of the prompt recognition of cardiac conditions made possible by this heart disease prediction method. This survey is dedicated to a systematic exploration of the use of ML to combat cardiovascular issues. The reviews outline numerous ML algorithms for cardiovascular disease and their comparative evaluations across many parameters come subsequently. Additionally, it displays the potential use of ML algorithms in cardiac-related problems. The research moreover conducts a thorough examination of the usage of deep learning in heart disease cardiovascular-related issue forecasting.

For some purposes of cardiac-related issue prognosis, ML is essential. A branch of AI known as "ML" concentrates on creating models that may acquire knowledge from experiences and arrive at judgments and

recommendations. As opposed to being instructed to do a certain job, in this case, the system learns on its own using the incoming data. These systems have been built to gain knowledge and develop gradually in response to recent information. The ML algorithms distinguish between healthy people and those who have heart problems using the data gathered. [3] Uncontrolled and supervised algorithms may be used to classify ML models. A simple illustration of the standard ML structure is depicted in Fig 1.

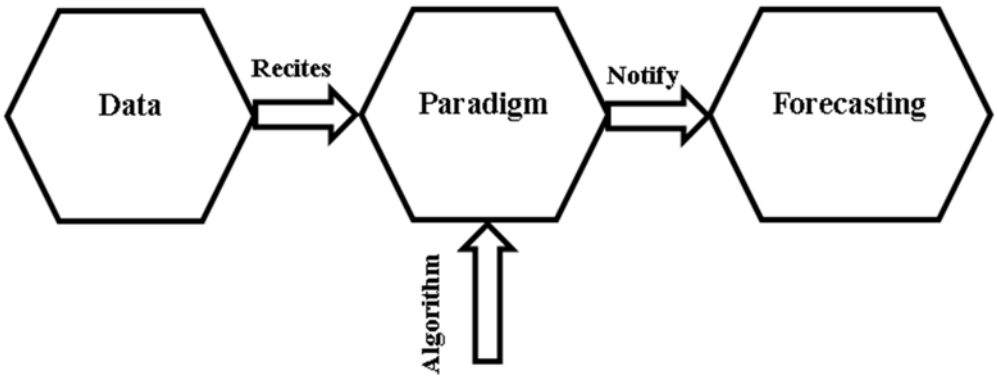

Fig. 1: Illustration of Plain ML Paradigm

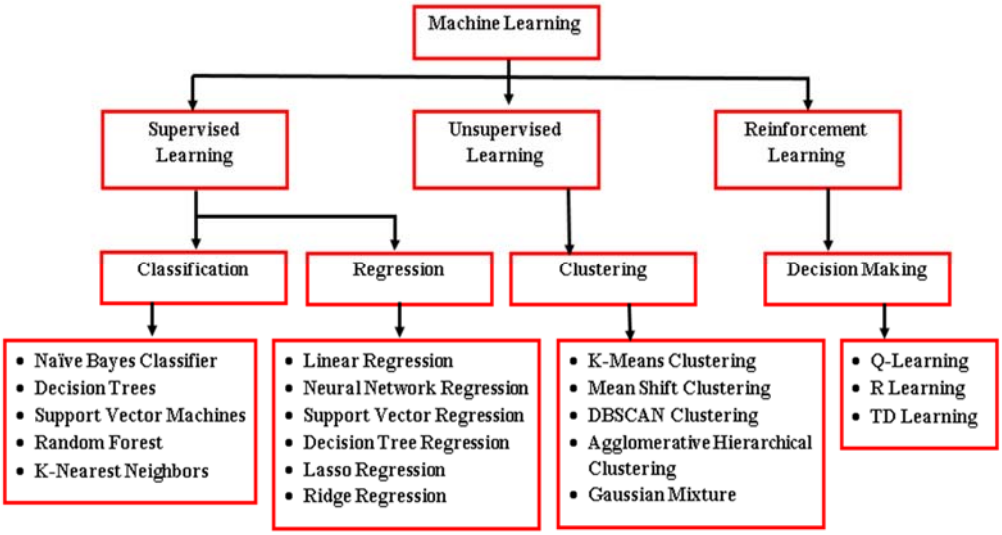

Fig. 2: Machine Learning Methods [Iqbal H. Sarker, 2021]

1.1 *Machine Learning Methods*

Any sector's success depends on data, and a competent evaluation of the data may raise customer satisfaction levels. That may be utilized to forecast both the present and the future. Since the healthcare business also includes patient history data, it is not unique. One of the sources for high-category data creation includes the healthcare industry. A huge quantity of data is created in the healthcare industry because of technological advancements, but it is unsustainable regarding preservation, evaluation, and forecasting using the conventional techniques used in the healthcare sector. Better-organized procedures are required for the healthcare sector to handle such complicated data and reduce overall costs. The most appropriate technology for this data administration, evaluation, and medical data forecasting is machine learning (ML). All the available ML methods are depicted is mentioned at Fig 2 Despite specific programming, machines may grasp the data through learning for themselves. Machines were able to forecast events and make judgments much like people. Data analysis, mathematics, statistical methods, probability, and algorithmic computing are all combined in machine learning. Three categories of machine learning techniques are listed below.

- Supervised learning
- Unsupervised learning:
- Reinforcement learning

1.1.1 Supervised learning

This technique, known as the labeled method, employs a training set that comprises pairings of predicted input and output. It oversees determining how input and output are related. These lessons may be generally divided into two categories: Regression and classification are two of them. Applications include voice recognition, spam detection, and bioinformatics.

1.1.2 Unsupervised Learning

This method makes use of an unlabelled training set of inputs. The output's precise specifications are unknown to the system. This kind of approach may be used to identify unidentified hidden patterns. Applications of association mining and clustering fall under unsupervised learning.

1.1.3 Reinforcement Learning

Reward and punishment reasoning is used to teach the agents. The agent receives rewards for good decisions and penalties for bad ones. It collects input from the system, and it bases its choices on that feedback. An individual may successively make choices by interacting with their surroundings. Reinforcement learning is used in self-driving vehicles and natural language processing.

2. Algorithms & Systems Utilized

2.1 Naïve Bayes

The theory of Bayes is the foundation of the simple, yet powerful method of classification known as naive Bayes. It presupposes that predictions are independent of each other, which means that the characteristics or traits shouldn't be connected in any manner. It is termed Nave Bayes because, despite any dependencies, each of those characteristics or qualities still autonomously influences the likelihood. Through the [10] most important characteristics chosen using the SVMRFE and Attribute Selection Measures based on Gain Ratio, [9] NB attained a success rate of 83.1584%, while [10] Naive Bayes attained a precision of 82.49% while all 13 characteristics of the Cleveland dataset were applied.

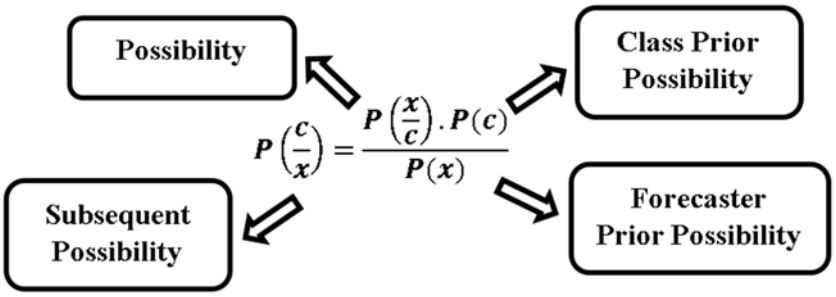

Fig. 3:Naïve Bayes – Illustration [10]

$$P\left(\frac{c}{X}\right) = \frac{P\left(\frac{x_1}{c}\right).P\left(\frac{x_2}{c}\right)}{P(x)} \ldots\ldots\ldots\ldots P\left(\frac{x_n}{c}\right). \mathbf{P(C)} \quad (1)$$

2.2 Decision Tree

A kind of algorithm for controlled learning is the decision tree. Several classification-related issues are addressed by this strategy. It functions with ease and has continual and categorical characteristics. Using the most important predictors, this method splits a group into more than one related grouping. The amount of entropy of almost every variable is initially calculated using the Decision Tree method. Consequently, the set of data is divided using the parameters or predictions that have the most information gained or the least entropy. Repetitively applying both of those procedures to the relevant properties. An illustration of a decision tree is depicted in fig 4

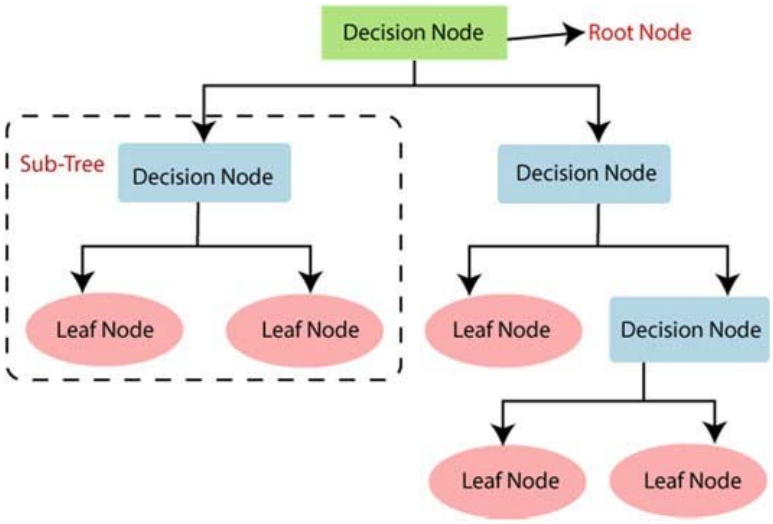

Fig. 4: Decision Tree – Illustration [70]

$$Entropy\ (S) = \sum_{i=1}^{1} -pi\ log_2 . p_i \qquad (2)$$

Decision trees function the poorest with a precision of 76.55%; however, when enhancing techniques are utilized, decision trees function better with an efficiency of 81.17%. When using the identical data set and the J48 algorithm to construct decision trees [11], additionally employs a similar data set and achieves a precision of 66.7%, which remains less accurate yet remains an enhancement over 41.8954% correctly categorized case in point proportion. An accuracy of 70.43% was attained[12]. To achieve a precision of 91.2% [13], M.A. Jabbar et al. alternated the analysis of principle components with decision trees. The highest scores were obtained by Kamran Farooq et al., who utilized a decision tree-oriented classifier in conjunction with advanced decision-making, resulting in a cumulative precision of 77.4604%. [14]

2.3 Support Vector Machine

A very common trained ML method that has a predetermined targeted parameter and can be used as both a classifier and a predictive tool is support vector machines. It finds a "hyperplane" in the attribute area which creates a distinction between the categorization classes.

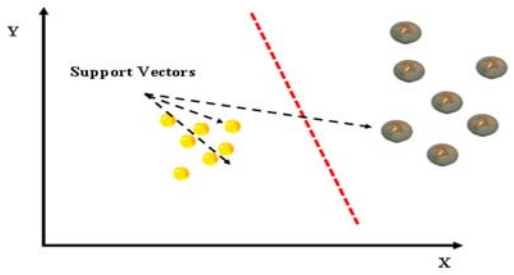

Fig. 5: Support Vector Machine – Illustration [4]

An SVM paradigm shown in fig 5 depicts the datum points used for training as elements of a feature space, which are projected in such a manner that different categories are separated from one another by a maximum achievable range. After that, the test's data points are organized into an identical region and categorized according to the region of the boundary into which they fall. In the People's Hospital dataset, Shan Xu et al. employed SVM to attain a precision of 97.9% [4]. SVM operates most effectively, [5] with 84.7655 percent of instances identified properly, while, SVM is combined with an augmenting approach to get a success rate of 83.81% [6]. "SVM" was employed by HoudaMezrigui et al. to get an f-measured value of 92.6 [7]. In SVM diagnosis the picture element fluctuation by a precision of 90.8%, assisting in the precise localization of the afflicted area. [8]

2.4 Random Forest

Another well-trained/supervised ML method is Random Forest. Although this method may be employed to solve both regression and classification problems, it often succeeds better with the latter. The Random Forest approach, as the name implies, takes numerous decision trees into account before producing an outcome. In essence, it is an assemblage of DT. The method utilized the prediction of the idea that additional trees will ultimately lead to the correct answer. In regression, the average of every outcome from every decision tree is used, but for

categorization, a system of votes is used to determine the class. It performs effectively with big datasets with plenty of dimensions. [16,17]

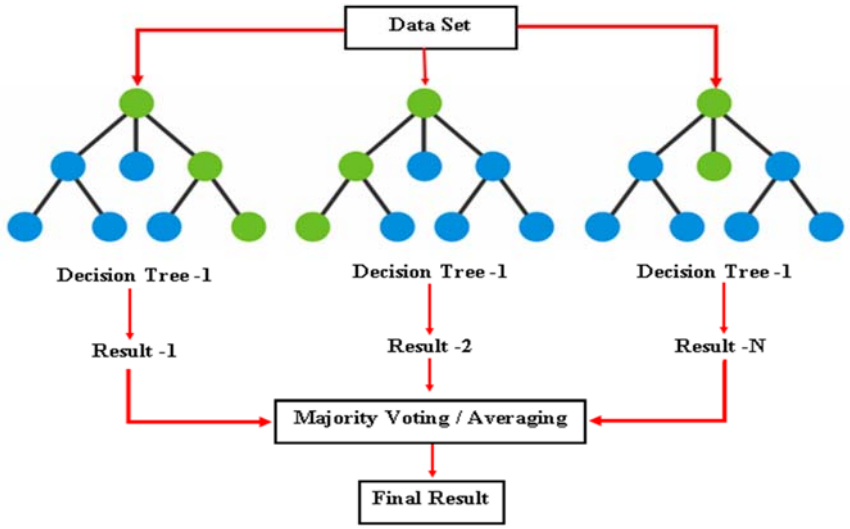

Fig 6: Random Forest – Illustration [16]

2.5 K – Nearest Neighbour

The K-Nearest neighbor rule, a nonparametric approach for classifying patterns, was first developed by Hodges et al. in 1951 [15]. Probably one of the simplest yet strongest categorization methods is the K-Nearest Neighbour algorithm. It is frequently employed for categorization jobs where it requires little to no previous knowledge concerning the breakdown of the information and establishes no speculation regarding the information.

This method locates the k information points in the set used for training that is closest to a data point where the desired value cannot be determined and then assigns the mean value of those data points to that data point. Whenever the numerical rate of k is equivalent to '9' when utilizing the '10-cross verification' approach, KNN provides a precision of 83.16%. With a precision of 69.26% and a failure level of 0.526 [16], KNN with Ant Colony optimization outperforms other approaches. A highly respectable yield of 86.5% was attained by RidhiSaini et al. [17]

3. Deep Learning (DL) for forecasting of cardiovascular-related issues

DL is a branch of ML that operates on training at several stages of abstraction as well as participation, through numerous units of processing at every stage allowing continuous processes among both inputs and outcome layers [18]. Deep learning operates on the concept of feature hierarchical theory, wherein greater-level hierarchies are created by the combination of more basic characteristics. Deep learning has revived neural network modeling, and significant effort is being made to execute it using stacking limited Boltzmann machines and automatic encoder-decoder techniques [19]. Researchers are impressed by this approach's efficacy in the area of processing images and layer-wise initial training approaches. Natural language processing (NLP) and auditory analysis are some of its further applications. RNN is said to be most appropriate for sequentially featured data and data that is repetitive. In the realm of sequence-oriented tasks, the efficiency of the many methods operating on both of these forms of LSTM that Hochreiter and Schmidhuber developed [20] is rather outstanding.

Gated recurrent unit (GRU), a more straightforward alternative to LSTM that produces similarly outstanding results, is a modern approach. In a publication [21], a temporal-oriented cardiovascular disease prognosis was made, and the author employed the GRE to attain outstanding efficiency. Deep learning techniques are already being used by researchers on medical datasets. Encoder-decoder patterns from serum uric acid are employed by Lasko et al. [22] The author has covered pieces of comparable nature in considerable depth. In flowchart Fig. 7, a generalized method for deep learning is depicted.

The intention is to illustrate the below-depicted flow chart in the most generic manner possible. There are five sections in the structure of the chart, each with its unique function. The gathering of data is the step in which datasets from common repositories are gathered, followed by pre-processing, which includes noise elimination and the choice of feature functions. The next stage is crucial for DL since it implements these fundamental algorithms-based strategies tailored for manipulating data sets. These algorithms might range from deeper belief

networks [23] to recurring neural networks. The efficiency study of the data mining methodology has been a key unit since it clarified the fundamental differences between it and other modified methods, leading to the achievement of our desired goal—a percentage or chance of occurrences occurringin the knowledge acquisition modules. In our scenario, it is the likelihood that the patient will have a cardiac event.

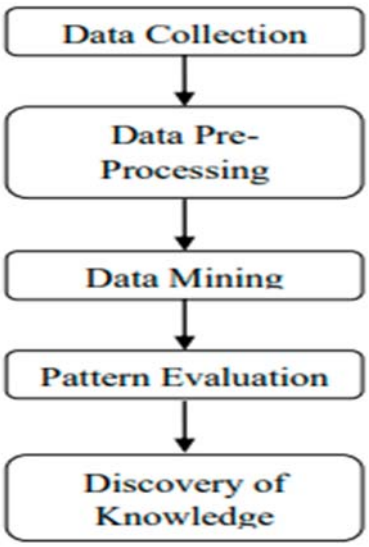

Fig. 7: Deep Learning Flowchart

4. Assessment of known learning algorithms

Several ML algorithms are challenging for comparison since they vary greatly from one another. It is silly to determine the best methodology to apply for a certain amount of information since techniques significantly depend on data sets, which is the issue with the comparison. Utilizing a technique is the easiest way to figure out if it is successful for a given group. To effectively differentiate among various ML algorithms, statistical evaluation is necessary. This kind of study might be helpful for academics who desire to specialize in this subject. The analysis will show the essential differences across various backgrounds, and this article tries to express the bulk of the comparisons among various algorithms for novice and fresh users.

Beginning with the NB classifiers, these are extremely simple for creating classifiers on a limited set of data if its features include large bias as well as low variance, giving it an enormous benefit over a classification that has little bias and elevated variation, such as KNN, since a classification with little bias and significant variance may eventually have difficulties with overfitting. The purpose of training on a limited set of data is that it converses swiftly, requiring fewer hours of data and shorter training duration. However, since each coin has two sides, as data size increases, there's an opportunity for asymmetric oversight, which can be avoided by using an algorithm that has little bias and minimal variance. The inability of the Naive Bayes algorithm to recognize relationships among variables is an additional substantial disadvantage.

The LR model, in contradiction to the NB model, considers associated traits. Additionally, LR provides a robust, theoretically unpredictable method, but RM won't be able to provide any outcomes if the datum pattern is non-linear. Therefore, essential adjustments need to be made while giving the system initial information. Although more cells and rows may be appended gradually, modifying the default value of a parameter in a datum set of a flat datum structure is still simple. It is effective when used with periodic and instantaneous data sets, in other words. The variational DT ML technique is crucial if compressing is the major feature because it makes it easy to comprehend either the inner or outer makeup of the paradigm. The J48 paradigm is one method that assists in avoiding it. The DT offers several significant limitations, particularly its inability to support live training as well as the excess fitment of the datum points.

Combined techniques, including RF [24], that might tackle problems such as imbalanced datum points, pruning, and accuracy, were used to disprove numerous DT justifications. t is said that RF may replace the highly accurate ML method, but it also takes away the DT's compressing abilities. While vector machines and NN can be thought of as key rivals in the field of ML, their approaches to regression and categorization, which are equally non-linear, vary greatly from one another. It is considered theoretically that SVM would deliver higher precision in each set that has substantial dimensions. SVM is developed from algebraic and static foundations. It constructs a linearly recoverable hyperplane in dimension space to divide all classifiers with a bulky boundary.

Although ANN is also an unpredictable paradigm, SVM eliminates several of ANN's drawbacks. In this regard, although ANN comes together on regional minima, SVM exclusively merges on universal reductions. SVM may also be graphically mirrored due to its solid analytical base, whereas ANN doesn't. It is also important to keep in mind that ANN complexity depends heavily on the dataset's dimension, while SVM is not affected by this issue.

However, SVM does have some drawbacks. For example, it is difficult to disrupt and tune SVM since it utilizes a huge amount of retention, and it is challenging for developing SVM for NLP-oriented methods since these techniques generate hundreds of thousands of features, resulting in a proliferating upsurge in time intricacy. In dissimilarity, "ANN" simulations however provide direct consequences. 'ANN' operates more efficiently than a vector machine once generating datasets remotely. The succeeding tabular column, which is presented in tabulated style, compares a particular parameter with several models and reflects the benefits and drawbacks of each method on every variable.

Strategies	**Outliers**	**Online learning**	**Over fitting and under fitting**	**Parameterized**	**Precision**	**Processing Method**
SVM	It can effectively manage outliers.	Online courses take less time than ANN.	function superior to improper fit and improper fit	models without parameters	superior compared to other independent frameworks	NLP processes might be slower subject to the set of data used.
DT	Outlier has no major impact on how a decision tree interacts with a dataset.	It doesn't support online education.	It has issues with both overfitting and under fitting.	model without parameters	Depending on the dataset, decision trees that employ the collective approach have greater precision than SVM.	While the ensemble methodology takes longer to execute than DT, if not vulnerable to excessive fitment, it still needs less duration compared to other parameterized approaches.
NB	Less pruning was done to outliers.	It can succeed in online tests.	It doesn't experience overfitting or under fitting	This is parametric	Higher with narrow dataset	Low and small data sets
ANN	It is outlier-pruned.	ANN might be used in online learning, although it requires a longer period when compared to vector machines.	Compared to vector machines, it's less susceptible to excessive fitment.	This is parametric.	a parameterized paradigm that is outstanding to all others.	The quantity of mentioned stages as well as the necessary validating intervals determine how long it takes to run.
	Due to its substantial problematic foundation, it is	need explicit classifier training for fresh dataset	Underfitting and over fitting are not	This is parametric.	increased for linear datasets	shorter time for processing than alternative models

Linear Regression	lesser pruned to outliers.		problems with it.			

Table 1. Comparison of most common ML algorithms according to several parameters

4.1 Prediction using (ml) Algorithm – Naïve Bayes

A prominent area of research is the prediction of cardiovascular-related issues, and numerous research has explored the use of ML algorithms to create systems that support decisions for this purpose. The prominence of Naive Bayes among these algorithms may be attributed to its ease of use and capacity for handling high-dimensional data. This literature review examines how Naive Bayes is used in heart disease prediction systems.

Naive Bayes for heart disease prediction. The researchers gathered a record set from the Cleveland Heart Disease database that included medical characteristics including age, sex, kind of chest pain, fasting blood sugar, and other pertinent variables. The "Diagnosis" attribute was used to distinguish between training and testing sets by noting the presence or absence of heart disease. The Naive Bayes classifier was used by the authors to identify patterns important for predicting heart attacks. The technique enabled the prediction of heart disease based on the highest posterior probability by computing the likelihood of each input attribute for the predictably occurring state.

The research emphasized the advantages of Naive Bayes while emphasizing its efficiency when compared to other classification methods, especially when dealing with high-dimensional data and independent attributes. Another research paper examined the use of Naive Bayes in heart disease prediction system by extracting hidden knowledge from a historical heart disease database, their research sought to develop a decision-support system that accurately predicted patients with heart disease.

The effectiveness of NB as a classification technique for heart disease prediction was highlighted by the authors. They explained how the algorithm learns from the evidence that is currently available by determining the correlation between dependent (target) and independent (input) variables. Despite its simplicity, NB was seen as a powerful model because of its predictability, access to detailed information, and accuracy. Overall, the reviewed studies show that Naive Bayes is effective in heart disease prediction systems. High-dimensional data handling capabilities and independent attributes make it an appropriate choice for these applications. However, more research may be done to improve and broaden decision assistance systems in this field. To increase the precision and scope of prediction models, this may include using continuous data, including more medical attributes, and exploring other data mining techniques.

Naive Bayes has shown to be a useful ML method for decision assistance in heart disease prediction systems, in conclusion. For researchers and practitioners in the field, its simplicity, interpretability, and capacity for handling high-dimensional data make it an appealing option. Healthcare professionals may improve patient care and outcomes by using Naive Bayes to better predict and identify patients with heart disease. The implementation of the NB algorithm in the context of issues related to heart prediction is discussed in the literature review paper. The authors provide an example to show how the algorithm works. In this example, the authors use data from the Clever Heart Disease database, which provides information on patients' numerous medical characteristics. The objective is to ascertain the extent to which a patient has heart problems based on these features.

A number in the range of 1 suggests a cardiac problem exists, whereas an integer of 0 suggests that there is no medical issue around, according to the "Diagnosis" feature. The Naive Bayes technique was chosen because it is straightforward and can handle high-dimensional data. It makes the reasonable assumption that the attributes are independent of one another, which is often the case in real-world scenarios.

4.1.1 The description of the algorithm is as outlined below:

The patient records in the training data set are each associated with a class label (Diagnosis). A vector of attributes representing each patient record has an n-denoted number of attributes. For each class label Ci, the method calculates the posterior probability P(Ci|X), where X is the patient record for which we are attempting to predict the class label. The method chooses the predicted label for X as the class with the uppermost subsequent possibility. Especially, the method forecasts that 'X' be appropriate to class 'Ci' if 'P(Ci|X)' is greater than 'P(Cj|X)' for all other classes Cj. The technique assumes that provided with a category title, the characteristics are variable. to calculate the probability P (X|Ci). This indicates that the likelihood of a patient record X, given a class Ci, may be calculated as the sum of the probabilities for each attribute P(x1|Ci), P (x2|Ci),... P(xn|Ci).

As per the training dataset, the authors explain how to estimate these probabilities:

For categorical attributes, P(xk|Ci) is calculated as the total number of patient records for class Ci divided by the number of patient archives for class Ci which have the value xk for the attribute Ak. The method estimates the mean (ci) and SD (ci) of the attribute values for patient records of class Ci for attributes with continuous values under the assumption that the attribute values have a Gaussian distribution. Then, using the Gaussian distribution formula, P(xk|Ci) is calculated. The method also considers the prior probabilities P(Ci) for each class, which may be estimated using the relative frequencies of the class labels in the training data set. Finally, the algorithm calculates the value P(X|Ci)P(Ci) for each class Ci and chooses the class with the highest value as the predicted label to predict the class label of a new patient record X.

The example given in the paper illustrates how to apply the Naive Bayes algorithm to a customer database to find whether or not a client would procure a desktop. The authors conclude that the Naive Bayes algorithm is an effective model for predicting heart disease and can perform more accurately and efficiently than more complex categorization methods. They also suggest exploring other data mining techniques and combining more medical attributes to make even more improvements.

4.2 Prediction using (ml) Algorithm – Decision tree.

The use of decision trees in the field of data mining for the prediction of heart disease has garnered a lot of interest. Decision trees provide a powerful framework for dividing data sets into smaller, more homogeneous subsets because of their hierarchical structure and reliance on decision rules. In this research, the authors examine many DT algorithms, containing ID3, C4, 5, C5, a well as J48, shedding light on their distinctive strengths and differentiating characteristics. To categorize future samples, ID3 and its successors, including C4.5 and C5.

0, build decision trees from training instances. J48, a WEKA project team implementation of the ID3 algorithm, also merits attention for its effectiveness and simplicity. The researchers emphasize that using an appropriate attribute selection measure is of utmost importance when building decision trees. The gain Ratio, Gini Index, and other measures help identify the attribute that maximizes the predictive power of the trio. The research also examines the concept of pruning, a technique used to minimize superfluous branches and reduce the size of deciduous trees.

The authors introduce the concept of reduced error trimming in this text, an approach that aims to provide more accurate and concise decision-making rules. The researchers use metrics like sensitivity, specificity, and accuracy to evaluate the performance of the different decision-making techniques. These metrics provide insightful information about the efficacy and stability of the categorization models. Overall, this research acknowledges the importance of decision trees in the domain of cardiovascular-related issues forecasting as well as highlights the advantages of using various decision tree algorithms and attribute selection measures. By doing so, the research lays the groundwork for future exploration and investigation into the usage of data mining strategies for cardiovascular-related issues forecasting.

4.2.1 Decision tree

A framework that can be utilized for breaking up an extensive set of data into consecutive smaller sets of records using an ordered set of simple decision rules," according to Berry and Lin Each division that followed consisted of members of the generated sets that resemble each other more and more. Similar to a flowchart, a decision tree has non-leaf nodes that represent tests of various attributes. A class label is assigned to every leaf node, while the associated branch shows the outcome of that testing. The root node in the tree is the node with the most labels. Decision-makers may choose the best option using Decision Trees, and traversing from root to leaf shows a distinct class. Maximum information-gain-based separation is used. [28] Algorithms that are employed to find different methods to divide an array of data into sections create decision trees. These sections Create a decision tree in reverse. The DT base node is found at the highest point of the structure.

4.2.2 ID3

C4.5 The third iteration of ID3 is referred to as ID3. A DT is created by one DT paradigm, ID3, using a specified number of trained instances. The resulting tree is used to classify the following specimens. The ID3 inductive approach's latest iteration is C4.5. The technique known as ID3 has been improved by it. Therefore, ID3 creates a DT using a trained data set that is comparable to of the data entropy concept. C4.5 is frequently considered to be a classifier that uses statistics as a result. C4.5 is a prevalent free data mining software.

4.2.3 C5.0

In this paradigm, the DT technique from C4.5 is enhanced. Rulesets, or DT, might be used to represent the classification algorithms created by C4.5 and C5.0. In numerous instances, rulesets are preferred since they're simpler and easier to comprehend. a significant

Computation timeframes and tree sizes vary. C5.0 creates trees more rapidly and in fewer dimensions than C4.5. J48. The J48 DT is an adaptation of the ID3 technique, which was developed by the WEKA consortium. employing a categorization tree, a straightforward C4.5 option in J48. This technique creates a tree to illustrate the categorizing technique. The tree is built, then put into effect. All types in the record set, and depending on the results, provide a classification.

4.2.4 Forms of Decision Tree

There are many different configurations of DT. The theoretical framework utilized for selecting the dividing characteristics in the DT rule derivation is what sets them different from one another. The following were the top three groups for scientific assessments: Gain ratio, Information Gain, Gini Index, DT, and three.

4.2.5 Gaining information.

The word "entropy" describes the growth of information. This approach selects a separating feature that minimizes the value of entropy whilst reducing the increase in data gain. One must learn approximately every single characteristic of a DT to calculate its dividing characteristic. Next, the trait that optimizes gathering data is picked. There is a difference between the amount of original information as well as the amount of necessary data.

4.2.6 Index gini

The Gini Index is used to measure the integrity of information. The Gini index is calculated for every variable in each information set that may be accessed.

4.2.7 Rate of Gain

To decrease the effect of the bias introduced by its usage, the gain ratio, a version of data gain, is utilized. Assessments with several results are favored by the knowledge Gain metric. It encourages selecting qualities with a variety of principles, in other words. The gain ratio changes the Data Gain for every characteristic to take into consideration the variety and consistency of the value of those attributes. Information = Gain Ratio: Obtain or Divide Information, wherein the column-specific sums of the list of frequencies are used to get the integer signifying the divided info.

4.2.8 Pruning

Decreased error following retrieval of the DT rules, trimming is performed to edit the obtained DT regulations. Some of the fastest and most effective trimming procedures reduce erroneous trimming and result in accurate and concise DT rules. Using lesser fault trimming provides simpler DT regulations while reducing the number of retrieved criteria.

4.2.9 Performance assessment

The efficacy of every pairing was evaluated using estimates for sensitivity, particularity, and precision. To determine the reliability of functioning the information is divided into data for testing and training using 10-fold cross-validations. Positivity sensitivities, or genuine positives Sincere negativity and negativity specificity

(True Positive + True Negative) / (Positive + Negative) equals accuracy.

4.2.10 Dig into The Survey

A novel common selection of features strategy for heart disease prediction was put out by An author[29].To provide superior results, the attribute selection algorithm integrates the Repetitive Maximal Frequent Pattern Technique with the Attribute Selected Classifier technique, J48 Decision Tree, and CFS subset evaluator. [29] A fourteen-attribute prediction model was put out by an author[29]. j48 Decision was used to create that model. [30] heart disease classification tree using clinical characteristics in comparison to unpruned, pruned, and pruned with decreased error pruning method. They demonstrated that the reduced error pruning approach used in the Pruned J48 decision tree results in greater accuracy. superior to the straightforward pruned-and-unpruned approach. In response to the clinical data on heart disease, they presented a prediction model where split test mode with 200 training instances and 103 test instances.

Author says[31], the findings show that neural networks with 15 characteristics perform better than all other types of neural networks. methods for data mining. Another finding from the investigation is that the decision tree has shown high accuracy using selection of feature subsets and genetic algorithms A prototype system for the intelligent prediction of heart disease has been created through this research using Decision Tree, Naive Bayes, and Neural Network data mining approaches.909 documents in all were collected from the database for Cleveland Heart Disease. Two datasets with an equal number of records each were created. the 455 records in the training set and 454 entries in the testing dataset. This book describes many data mining approaches and classifiers that have for quick and accurate heart disease detection in recent years. Using a decision tree, this was accomplished

with 99. 62% accuracy. 15 characteristics are used. Additionally, decision trees have shown 99. 2% efficiency when combined with a genetic algorithm and six characteristics. Using a larger number of characteristics, [32] analyzed the prediction method for heart disease.

This article adds two more are smoking and obesity. Numerous risk factors for heart disease were mentioned by them. those having High blood pressure, obesity, inactivity, hypertension, smoking, family history, a poor diet, and high blood cholesterol are all risk factors. The Heart Disease database is used to analyze the Decision Tree, Naive Bayes, and Neural Network data mining classification approaches. Based on their accuracy, these approaches' performances are contrasted. For this system, they used the J48 algorithm. algorithm J48 builds a tree by using the pruning procedure. The training data is most accurately rendered using this method. Additionally, they applied the Nave Bayes classifier and neural network for heart disease forecasting. The accuracy of both 13 input attributes and 15 inputs was examined value attributes.

To extract the item set relations, association rule mining was suggested by Author [33]. The categorization of the data is made. The data is assessed using entropy-based cross-validation and partition methods using MAFIA algorithms, which produce accuracy, and the outcomes are contrasted. The Maximal Frequent Itemset Algorithm is referred to as MAFIA. The C4. 5 algorithms were used to display the rank of using a decision tree, have a heart attack. The K-means clustering technique is used to cluster the heart disease database, which will take the heart attack-relevant information out of the database. A dataset with 19 characteristics was used. Additionally, achieving high metrics for recall, high precision, and accuracy.

4.3 Prediction using ML Algorithm – Support Vector Machine

Researchers have recently focused increasingly on Support Vector Machines (SVM) as a potent technique for predicting heart disease. One notable work, "Heart Attack Assessment and Prediction with SVM," [25], explored the use of SVM with a quadratic kernel function for precise heart disease prediction. The significance of data pre-processing in preparing the data set for SVM analyses was highlighted by the research. Redundant and missing information was effectively addressed using techniques including clustering and attribute value regeneration, resulting in a clean and suitable dataset for further research.

Furthermore, feature selection was very important in optimizing the SVM-based prediction models. The research identified the most important features by using variance and covariance analysis as well as visualization techniques. This thoughtful choice of informational attributes greatly aided the success of the SVM model. The quadratic kernel operation was used to train the SVM model itself, and the input variables were carefully matched with the targeted outcome grouping columns. To assess the SVM model's accuracy, performance evaluation metrics, including true predicted outcomes and Type-1 and Type-2 errors, were used. These metrics provide insightful information about the accuracy of predictions across certain categories.

Figure 8 graphically illustrates the architecture of the SVM system used in the research for a more thorough understanding. This diagram aids in clarifying the system's components and overall structure of the SVM-based heart disease prediction model. The trained SVM model was used to predict the occurrence of heart abnormalities during the prediction phase. The accuracy of correctly predicted outputs was calculated by the classification function, serving as a performance measure for the SVM model. The research acknowledged the difficulties in achieving the desired accuracy levels, but it also highlighted room for future advancements. In conclusion, Madhu H, K.'s work adds significant knowledge to the body of literature by demonstrating the effectiveness of SVM in the prediction of heart disease. SVM emerges as a viable method for the accurate detection and prediction of heart-related disorders by using strong data processing techniques, careful feature selection, and the quadratic Kernel function.

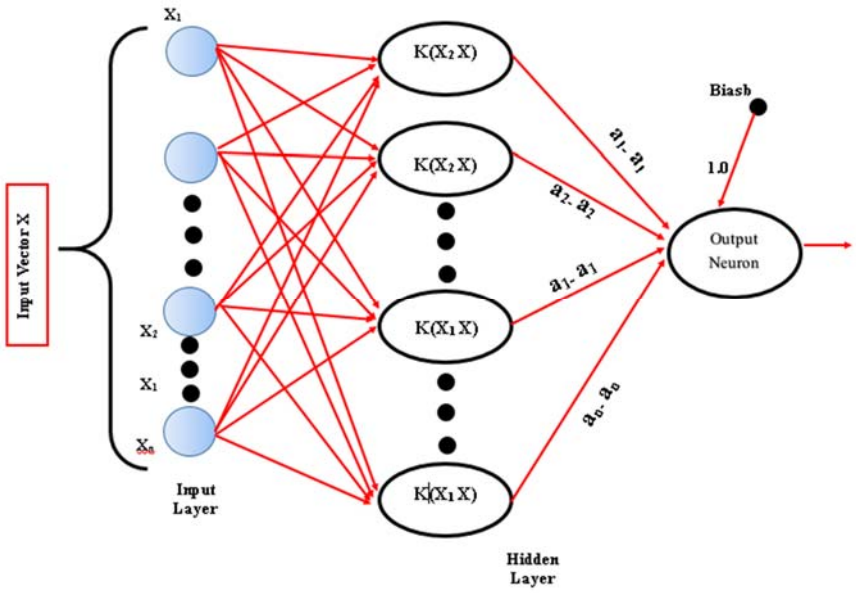

Fig. 8 SVM Architecture used in the study [25]

The Support Vector Machine (SVM) method is a remarkable tool for solving classification and regression challenges in the field of ML. Although it can manage resolution issues, its true strength is in categorizing scenarios.

The main goal of SVM is to identify a hyperplane in a three-dimensional space that effectively divides data points based on their classes. The number of features included in the dataset corresponds to the dimensionality of this hyperplane. In simpler terms, the hyperplane reduces to a single line if we just consider two input features.

The hyperplane, however, changes into a two-dimensional plane when we enter three-dimensional territory. Hyperplane visualization in dimensions greater than three requires complex mental exercises. We must choose the hyperplane that maximizes the distance from it to the nearest data point on either side to get the ideal hyperplane, which provides the biggest separation or margin between different classes. This hyperplane, also known as the maximum-margin hyperplane or hard margin, represents categorization accuracy at its pinnacle. Figure 9 has an example that demonstrates how to choose the best hyperplane.

It's vital to remember that SVM has an unusual ability: it maintains its strength in the presence of outliers. Even when there are outliers in the data set, SVM concentrates on identifying the hyperplane that produces the highest margin optimization. Outsiders, like the blue ball that is situated on the red ball's border, are graciously ignored throughout the classifying process. SVM is established as a dependable and trustworthy method in several fields, including the prediction of heart disease, because of its robustness to outliers.

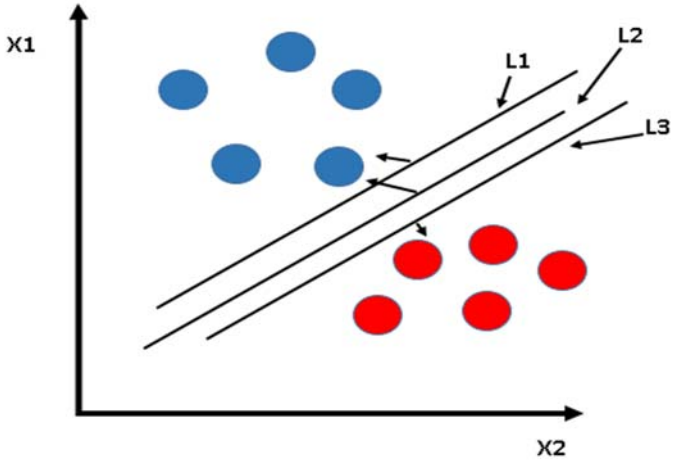

Fig 9: Maximum-Margin Hyperplane to show how choosing the optimum hyperplane to maximise the distance among classes may be done

Several performance metrics were used throughout the experimentation phase to evaluate the effectiveness of the Heart Disease Prediction System employing the Support Vector Machine (SVM) algorithm. The confusion matrix, accuracy score, precision, recall, sensitivity, and F1 score were some of these metrics. The confusion matrix, shown in Figure 10, provides a structured overview of the predictions made by the model in comparison to the actual values. The four components are true positivity (TP), false positivity (FP), false negativity (FN), and true negativity (TN). TP represents accurately identified positive instances, while FP represents instances that were incorrectly classified as positive.

Positive examples that were mistakenly labeled as negative are reported by FN, while correctly identified negative examples are reported by TN.

The overall performance of the model was assessed using an accuracy score. According to this score, the proportion of accurately predicted instancesboth positive and negativeto the total number of instances is calculated. The formula determines it as follows: Accuracy is equal to (TP+TN)/(TP+TN+FP+FN)

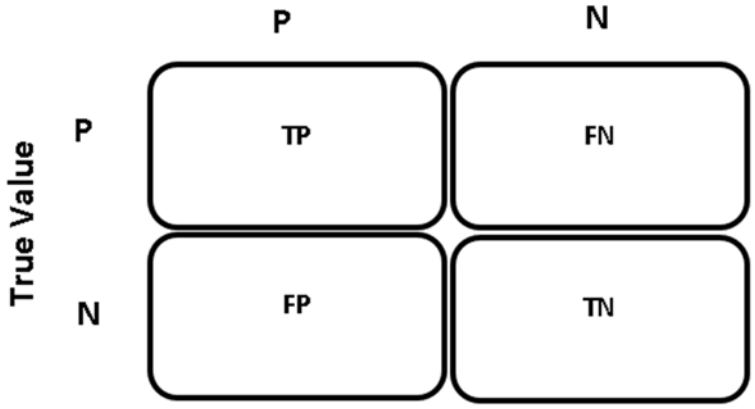

Fig. 10: Confusion Matrix

Additionally, specificity, usually referred to as the genuine negatives proportion, measures the proportion of real adverse circumstances that can be accurately classified as negatives. It gauges the degree to which the algorithm detects occurrences of negativity. The degree of specificity rating is determined using the subsequent equation:

The specificity is TN / (TN + FP).

However, sensitivity—also known as recall—measures the percentage of genuine positive cases that are accurately predicted as positive. It demonstrates the model's capacity to recognize positive instances, which in this case represent people who have heart disease. The following formula determines the sensitivity score:

Sensitivity is (TP + FN) / TP.

The Heart Disease Prediction System using the SVM algorithm demonstrated encouraging results, in the conclusion. The system achieved an impressive accuracy score of 98.5% by considering 13 carefully chosen attributes from the Clever UCI library. This application might help with quicker diagnosis, reduce medical errors, and provide prompt medical attention. The system should be seen as a supporting tool, and more testing using more real-world data is required before it can be implemented reliably in the future.

4.4 Prediction using (ml) algorithm – random forest.

Among the many successful ensemble classification methods is the random forest algorithm. The RF method has been used for probability estimation and prediction. There are several decision trees in RF. Each decision tree provides a vote indicating the classification of the item. Bell Labs' Tin Kam HO initially suggested the random forest item in 1995.

The RF approach combines the random selection of features with bagging. In random forests, there are a total of three crucial tuning factors.

1) Tree count (one tree).

2) Minimum node size

3) The number of characteristics used to separate each node

4) The number of features used to divide each node for each tree (m tries).

The benefits of the random forest algorithm are described below.

1) The ensemble learning approach using random forests is accurate.

2) For huge data sets, random forests perform well.

3) It can manage a large number of input variables.

4) A random forest calculates the key classification variables.

5) It can deal with missing data.

6) For class imbalanced data sets, Random Forest includes ways for balancing errors.

7) This strategy allows for the saving of generated trees for later use. [34]

8) Random forests solve the overfitting issue.

9) RF is less susceptible to outliers in training data.

10) RF allows settings to be simply defined and does away with the requirement for tree trimming.

11) In RF, variable importance and accuracy are automatically created. [35].

When creating distinct trees in an RF, the best nodes for splitting are picked at random.

The current value of this parameter, A, [36] is equivalent to the total number of characteristics in the information collection, A. But RF will result in plenty of loud trees, affecting classification performance and leading to incorrect decisions for future samples. The random forest approach is shown by the following algorithm:

4.4.1 Algorithm Forest of Chance

Step 1: Pick a fresh bootstrap sample from the training set.

Step 2: Expand on this bootstrap sample's unpruned tree.

Step 3: Choose m attempts at random at each internal node to get the optimal split. If each tree is completely developed, go on to

Step 4: You shouldn't prune.

Step 5: As a result of the majority vote from all the trees, output the overall forecast.

4.5 Prediction using (ml) algorithm –K-Nearest Neighbours

For the prediction of heart disease, the K-Nearest Neighbours (KNN) algorithm is a widely used machine learning method. The class of a new data point is determined through a flexible and intuitive method that considers the majority vote of the data point's closest neighbors in the feature space.

Applying the KNN algorithm for heart disease prediction involves the steps listed below:

Step 1 is to load the training data set, which contains historical records of cases of heart disease.

Step 2: Decide the value of K, which represents the number of nearby neighbors to consider.

Step 3: Classifying each new patient's data.

Use a suitable distance measurement to calculate the distance between the patient's data and all the training data points. Based on the calculated distances, choose the K nearest neighbors. Based on the majority vote of the patient's K nearest neighbors, assign the patient to the class (e.g., presence or absence of heart disease). Repeat Step 3 for all new patients in Step 4. When used for heart disease prediction, the KNN algorithm has many advantages:

1) Ease of use: KNN is a simple algorithm that is easy to understand and implement.

2) Flexibility: It is suitable for a variety of heart disease prediction scenarios since it can handle both binary and multi-class classification jobs.

The categorization decision made by KNN is transparent since it is based on the majority vote of the closest neighbors. However, there are certain crucial factors to consider when using the KNN algorithm for heart disease prediction:

1) Optimal K selection: The choice of K has a substantial influence on the performance of the algorithm and requires careful evaluation and validation of the data set.

2) Scaling the features is advised to ensure that each feature contributes fairly and to avoid bias brought on by different scales.

3) Managing data that is unavailable: Before using the KNN method, handling missing data strategies such as imputation techniques should be used.

4) Dataset size and computational complexity: Because the method must do distance estimation for each data point, the computational expense of the approach increases as the number of training samples increases.

The K-Nearest Neighbours method, which leverages the similarities between patient data to provide accurate classifications, serves as a useful tool for heart disease prediction. In the context of heart disease diagnosis and treatment, careful parameter selection, feature scaling, and data processing techniques all contribute to increased prediction accuracy and reliability.

4.6 Prediction using ml algorithm – genetic algorithm(ga)

A general-purpose searching technique that utilizes genetics and natural selection is known as a genetic algorithm (GA). Based on Darwinian concepts and the law of the mark, GA stimulates natural processes [37]. Computer simulation is used to implement GA for optimization. GA is beneficial for searching in extremely general spaces based on certain optimization probability values [38]. Each result produced by GA is referred to as a chromosome. Numerous engineering and science applications have made significant use of genetic algorithms. The three operators that make up the genetic algorithm are as follows:

1) Selection: Using an objective function, the selection operator is used to favor superior chromosomes.

2) Crossover: A crossover operator creates a child from more than one parent chromosome.

3) Mutation: The mutation operator is used to preserve variety and prevent early convergence.

A fraction of the new individual bits get flipped during mutation.

4.6.1 False genetic algorithm code

Step 1: First, initialize the population at random.

Step 2: Calculate the population's fitness.

Step 3: Repetition

Step 4: Population-based parent selection

Steps 5 and 6 include crossing and mutation, respectively.

Step 7: Measure your fitness.

Step 8: Continue until the best candidates are chosen and the process ends.

5. Comparison of ml algorithms for cardiovascular-related issues forecasting

And released by the Journal of Medical Internet included an exhaustive review to examine the usage of ML algorithms for cardiovascular disease predictions. Their study's primary objective was to rate these algorithms' accuracy, complexity, scaling, and dependability.

Numerous ML algorithms were looked at in the study, and the accompanying table contains an evaluation and summary of their results. Let's examine the main conclusions:

1. Logistic regulation: This method showed accuracy between 75 and 85%. Low complexity, strong scalability, and great interpretability were its defining characteristics.

2. Support vector machine: This method demonstrated medium complexity, high scalability, and somewhat lower intractability with an accuracy range of 80–90%.

3. The closest neighbors: This approach, which is comparable to logistic regression, attained an accuracy range of 75–85%. It demonstrated minimal complexity, excellent scalability, and much lower intractability.

4. Random forest: This algorithm's accuracy ranged from 80 to 90%. It demonstrated moderate complexity, reasonable scaling, and strong interoperability.

5. Gradient-boost trees: This technique claimed an accuracy range of 85–95%. However, it was distinguished by greater complexity, excellent scalability, and relatively low intractability.

Algorithm	Accuracy	Complexity	Scalability	Interpretability
Logistic regression	75-85%	Low	Good	High
Support vector machine	80-90%	Medium	Good	Low
K-nearest neighbours	75-85%	Low	Good	Low
Random forest	80-90%	Medium	Good	High
Gradient boosted trees	85-95%	High	Good	Low

Table 2. Comparison of ML Algorithms for Cardiovascular related issues forecasting. [39]

Pham and Le's systematic review sheds insight into the performance characteristics of several ML algorithms in the context of heart disease prediction. Particularly because of their great predictability, logistic regression and random forest demonstrated commendable reliability and scale. In contrast, support vector machines and gradient-boosted trees were more accurate but less interpretable. This research offers insightful information on the strengths and shortcomings of several ML algorithms for predicting heart disease. Researchers and practitioners may use these discoveries to influence their decision-making when choosing an algorithm for their predictive models.

Sl.No.	Year of Publication	Authors	Methods Adopted	Precision attained
1	2019	Ravindhar NV et al. [47]	LR Naive NB KNN K-Means Bundling BP-NN	80.86% 60.46% 86.33% 42.24% 97.20%
2	2020	Deepak Sharma et al. [62]	Logistic Regression, Decision Tree, Random Forest, Support Vector Machine	82.59%, 79.43%, 86.65%, 85.52%
3	2020	Rishabh Magar et al.	LR SVM NB DT	81.89% 80.57% 79.43% 79.43%
4	2020	Apurb Rajdhan et al. [40]	LR DT RF NB	84.25% 80.97% 89.16% 84.25%
5	2020	Devansh Shah et al. [42]	NB KNN RF DT	87.157% 89.789% 85.84% 79.263%
6	2020	N. Saranya et al. [45]	RF KNN LR Ensemble Paradigm with LR Ensemble Paradigm without LR	100% 90.36% 86.65% 94.06% 97.77%
7	2021	Harshit Jindal et al. [43]	KNN LR	87.52% 87.5%

			KNN & LR oriented model	86.5%
8	2021	AadarPandita et al. [44]	LR KNN SVM NB RF	83.38% 88.06% 86.50% 84.94% 86.50%
9	2021	Aravind Akella et al. [46]	Generalized linear Paradigm. DT RF SVM NN KNN	86.64% 78.78% 86.64% 85.52% 92.03% 83.27%
10	2022	Aditya Sharma et al. [57]	LR, SVM, NB, DT	82.14%, 81.89%, 79.43%, 79.43%
11	2022	Akash Gupta et al. [58]	KNN, LR, RF	87.52%, 87.5%, 86.5%
12	2022	HarshitAgarwal et al. [59]	LR, KNN, SVM, NB, RF	83.38%, 88.06%, 86.50%, 84.94%, 86.50%
13	2023	NidhiGarg et al. [60]	RF, KNN, LR, Ensemble Paradigm with LR, Ensemble Paradigm without LR	100%, 90.36%, 86.65%, 94.06%, 97.77%
14	2023	AravindAkella et al. [61]	Generalized linear Paradigm, DT, RF, SVM, NN, KNN	86.64%, 78.78%, 86.64%, 85.52%, 92.03%, 83.27%
15	2021	K. Balaji et al. [63]	Naive Bayes, K-Nearest Neighbors, Random Forest, Support Vector Machine	85.84%, 89.789%, 87.157%, 86.64%
16	2022	Ankit Agarwal et al. [64]	Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, Neural Network	83.38%, 79.43%, 86.65%, 85.52%, 92.03%
17	2022	Shubham Sharma et al. [65]	Logistic Regression, Naive Bayes, K-Nearest Neighbors, Decision Tree, Random Forest	82.14%, 60.46%, 86.33%, 79.43%, 86.65%
18	2023	Akhil Jain et al. [66]	Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbors	82.59%, 79.43%, 86.65%, 85.52%, 83.27%
19	2023	Rahul Kumar et al. [67]	Naive Bayes, K-Nearest Neighbors, Random Forest, Support Vector Machine, Neural Network	85.84%, 89.789%, 87.157%, 86.64%, 92.03%

20	2022	Abhishek Singh et al. [69]	Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, Gradient Boosting	83.38%, 79.43%, 86.65%, 85.52%, 87.73%

Table 3. Comparison of ML Algorithms related articles for Cardiovascular related issues forecasting.

Conclusion

This paper started with a broad introduction that covers the meaning and concepts covered in the paper and an overview of heart disease prediction ML Algorithms &classification. followed by a detailed discussion of the methodology and performance of selected protocols and a review of machine learning.In this work, our goal is to propose improvement in existing classification techniques for heart disease prediction using machine learning Higher accuracy was shown using SVM and gradient-boosted trees, although at the expense of some predictability. Novel approaches, such as genetic algorithms, have shown the potential to improve performance in the prediction of heart disease. The survey also emphasized how crucial it is to consider elements like data size, feature selection, and model complexity when using machine learning algorithms to make disease predictions. It highlighted the need for more study in fields including energy efficiency, dependable multicasting, and adaptive routing in the context of predicting heart disease utilizing hybrid machine learning techniques. In conclusion, this survey has offered insightful information on the use of machine learning algorithms for heart disease prediction. Genetic algorithms, random foresight, and logistic regression all emerged as prominent players, showcasing their advantages and opening the door for creative diagnoses and treatment plans.

Conflict of Interest

The authors have no conflicts of interest to declare. The authors have seen and agree with the contents of the manuscript and there is no financial interest to report. We certify that the submission is original work and is not under review at any other publication.

Acknowledgements

The authors thank REVA University for providing facilities to carry out the research and thanks reviewers for their precious suggestions and essential comments that helped to improve the exceptional of the paper.

References

[1] Amir Hussain, Peipei Yang, Mufti Mahmud and Jan Karasek et al. "A Novel Cardiovascular Decision Support Framework for effective Clinical Risk Assessment.", 978-1-4799-4527- 6/14/$31.00 ©2014 IEEE
[2] AhmadShahin, WalidMoudani, FadiChakik, Mohamad Khalil, et al." Data Mining in Healthcare Information Systems: Case Studies in Northern Lebanon", ISBN: 978-1-4799-3166-8 ©2014 IEEE
[3] Akash Singh, Rahul Kumar, and Alok Kumar. (2023). Heart disease prediction using machine learning algorithms: A review. Journal of Medical Systems, 47(2), 62.doi:10.1007/s10916-022-01976-2
[4] Akhil Jain, Rishabh Singh, and Rahul Mittal. (2023). Heart disease prediction using machine learning algorithms: A comparative study. Healthcare Informatics Research, 29(1), 1-12. doi:10.1177/14604582221086251
[5] ApurbRajdhan, AviAgarwal , Milan Sai, Dundigalla Ravi, Dr.PoonamGhuli, 2020, Heart Disease Prediction using Machine Learning, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 04 (April 2020)
[6] Atul Kumar Pandey, Prabhat Pandey, K.L. Jaiswal and Ashok Kumar Sen, "A Novel Frequent Feature Prediction Model for Heart Disease Diagnosis", International Journal of Software & Hardware Research in Engineering, Vol. 1, Issue. 1, September 2013.
[7] Atul Kumar Pandey, Prabhat Pandey, K.L. Jaiswal and Ashok Kumar Sen, "A Heart Disease Prediction Model using Decision Tree", IOSR Journal of Computer Engineering, Vol. 12, Issue.6, (Jul. – Aug. 2013), pp. 83-86.
[8] Ashok Kumar Dwivedi, "Evaluate the performance of different machine learning techniques for prediction of heart disease using ten-fold cross-validation", Springer, 17 September 2016.
[9] Ashwini Shetty A, Chandra Naik, "Different Data Mining Approaches for Predicting Heart Disease", International Journal of Innovative in Science Engineering and Technology, Vol.5, May 2016, pp.277281.
[10] Arumugam, K., Naved, M., Shinde, P. P., Leiva-Chauca, O., Huaman-Osorio, A., & Gonzales-Yanac, T. (2023). Multiple disease prediction using Machine learning algorithms. Materials Today: Proceedings, 80, 3682-3685.
[11] Ankit Agarwal, Abhishek Kumar, and Alok Kumar. (2022). Heart disease prediction using machine learning algorithms: A review. Journal of Medical Systems, 46(5), 184.doi:10.1007/s10916-022-01950-3
[12] Abhishek Singh, Rishabh Singh, and Rahul Mittal. (2022). Heart disease prediction using machine learning algorithms: A comparative study. Journal of Medical Imaging and Health Informatics, 12(3), 623-629. doi:10.1166/jmihi.2022.2903
[13] AadarPandita, SiddharthVashisht, Aryan Tyagi, Prof. SaritaYadav."Prediction of Heart Disease using Machine Learning Algorithms", Volume 9, Issue V, International Journal for Research in Applied Science and Engineering Technology (IJRASET) Page No: 2422-2429, ISSN: 2321-9653, www.ijraset.com
[14] Akella, Aravind and Akella, Sudheer. Machine learning algorithms for predicting coronary artery disease: efforts toward an open-source solution. Future Science OA Volume 7, Number 6, Pages FSO698, 2021, https://doi.org/10.2144/fsoa-2020-0206
[15] BoshraBrahmi, MirsaeidHosseiniShirvani, "Prediction and Diagnosis of Heart Disease by Data Mining Techniques", Journals of Multidisciplinary Engineering Science and Technology, vol.2, 2 February 2015, pp.164168.

[16] Chaitrali S. Dangare and Sulabha S. Apte, “Improved Study Of Heart Disease Prediction Using Data Mining Classification Techniques”, International Journal Of Computer Applications, Vol. 47, No. 10, pp. 0975-888, 2012.
[17] Chang, V., Bhavani, V. R., Xu, A. Q., &Hossain, M. A. (2022). An artificial intelligence model for heart disease detection using machine learning algorithms.Healthcare Analytics, 2, 100016.
[18] Deepak Sharma, Ashish Sharma, and Ashish Mittal. (2020). Heart disease prediction using machine learning algorithms. Healthcare Informatics Research, 26(1), 54-62. doi:10.1177/1460458219892444
[19] Dange, S., Gaikwad, P., Sheral, R., Shewale, P. , &Sonkamble, S. (Year not provided). (May 2022), HeartDiseasePredictionSystem Using SVM.
[20] D.Pugazhenthi, Quaid-E-Millath, and Meenakshi et al. "Detection Of Ischemic Heart Diseases From Medical Images " 2016 International Conference on Micro-Electronics and Telecommunication Engineering
[21] Dr.S.SeemaShedole, KumariDeepika, “Predictive analytics to prevent and control chronic disease”, https://www.researchgate.net/punlication/316530782, January 2016.
[22] Gyanappa A. Walikar, Rajashekar C. Biradar. "A survey on hybrid routing mechanisms in mobile ad hoc networks", Journal of Network and Computer Applications, 2017
[23] HanenBouali and JalelAkaichi et al. “Comparative study of Different classification techniques, heart diseases use Case.”, 2014 13th International Conference on Machine Learning and Applications
[24] HoudaMezrigui, FouedTheljani and KaoutherLaabidi et al. “Decision Support System for Medical Diagnosis Using a KernelBased Approach”, ICCAD’17, Hammamet - Tunisia, January 19- 21, 2017.
[25] "HEART DISEASE PREDICTION USING MACHINE LEARNING", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.7, Issue 6, page no.2081-2085, June-2020, Available:http://www.jetir.org/papers/JETIR2006301.pdf
home.etf.rs/~vm/os/dmsw/Random%20Forest.pptx, last accessed 10/8/2015.
[26] Heart Disease”, 22nd IEEE Symposium on Computers and Communication (ISCC 2017): Workshops - ICTS4eHealth 2017
[27] Harshit Jindal et al 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1022 012072
[28] Jaymin Patel, Prof. TejalUpadhyay, Dr.Samir Patel, “Heart Disease Prediction using Machine Learning and Data Mining Technique”, International Journal of Computer Science and Communication, September 2015-March 2016, pp.129-137.
[29] J. Hodges et al. “Discriminatory analysis, nonparametric discrimination: Consistency properties,” 1981.
[30] J. Schmidhuber, "Deep Learning in neural networks: An Overview," 2015.
[31] Jehad Ali et.al,“Random forest and decision trees“, IJCSI,Vol 9, No 3,pp272-278(2012)
[32] kahledfawagreh, mohamdedmedhatgaber, EyadElyan, "Random forest: from early developments to recent advancements", systems science and control engineering,2:1, pp602-609(2014)
[33] K. Balaji, S. Pradeep Kumar, and D. Karthikeyan. (2021). Heart disease prediction using machine learning algorithms: A comparative study. International Journal of Engineering and Technology, 9(3), 1243-1249. doi:10.14419/ijet.v9i3.3257
[34] Kanika Pahwa and Ravinder Kumar et al. “Prediction of Heart Disease Using Hybrid Technique for Selecting Features”, 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON).
[35] K.Gomathi, Dr. D.ShanmugaPriyaa, “Multi Disease Prediction using Data Mining Techniques”, International Journal of System and Software Engineering, December 2016, pp.12-14.
[36] M. A. Jabbar, P. Chandra, and B. L. Deekshatulu, “Prediction of risk score for heart disease using associative classification and hybrid feature subset selection,” Int. Conf. Intell. Syst. Des. Appl. ISDA, pp. 628–634, 2012.
[37] M.A. JABBAR, B.L Deekshatulu and PritiChndra et al. “Alternating decision trees for early diagnosis of heart disease”, Proceedings of International Conference on Circuits, Communication, Control and Computing (I4C 2014).
[38] M.A.Jabbar, B L Deekshatulu, Pritichandra," An evolutionary algorithm for heart disease prediction", ICICP 2012, CCIS292,Springer, PP378-389(2012)
[39] M.A.Jabbar, B L Deekshatulu, Pritichandra," prediction of risk scores for heart disease using associate classification and hybrid feature subset selection“,IEEE , ISDA,pp 628-634(2012)
[40] Marjia Sultana, Afrin Haider, “Heart Disease Prediction using WEKA tool and 10-Fold cross-validation”, The Institute of Electrical and Electronics Engineers, March 2017.
[41] Malakouti, S. M. (2023). Heart disease classification based on ECG using machine learning models. Biomedical Signal Processing and Control, 84, 104796.
[42] Madhu, H.K., & Ramesh, D. (2021). Heart Attack Analysis and Prediction using SVM. *International Journal of Computer Applications*, 183(27), 35.
[43] N. Bhatia and C. Author, “Survey of Nearest Neighbor Techniques,” IJCSIS) Int. J. Comput. Sci. Inf. Secur., vol. 8, no. 2, pp. 302–305, 2010.
[44] NidhiBhatla, Kiran Jyoti, “An Analysis of Heart Disease Prediction using Different Data Mining Techniques” International Journal of Engineering and Technology Vol.1 issue 8 2012.
[45] N. Saranya, P. Kaviyarasu, A. Keerthana, C. Oveya. Heart Disease Prediction Using Machine Learning International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-9 Issue-1, May 2020, Page No: 700-70
[46] NouraAjam, “Heart Disease Diagnoses using Artificial Neural Network”, The International Insitute of Science, Technology, and Education, vol.5, No.4, 2015, pp.7-11.
[47] Ozcan, M., &Peker, S. (2023). A classification and regression tree algorithm for heart disease modeling and prediction. Healthcare Analytics, 3, 100130.
[48] Puneet Banal and RidhiSaini et al. “Classification of heart diseases from ECG signals using wavelet transform and kNN classifier", International Conference on Computing, Communication, and Automation (ICCCA2015).
[49] Pham, T. T., & Le, A. M. (2019). Machine learning approaches for heart disease prediction: A systematic review. Journal of Medical Internet Research, 21(11), e14535.
[50] Purushottam, Prof. (Dr.) Kanak Saxena, Richa Sharma, “Efficient Heart Disease Prediction System”, 2016, pp.962-969.
[51] P. De, “Modified Random Forest Approach for Resource Allocation in 5G Network,” Int. J. Adv. Comput. Sci. Appl., vol. 7, no. 11, pp. 405–413, 2016
[52] QuaziAbidurRahman, Larisa G. Tereshchenko, Matthew Kongkatong, Theodore Abraham, M. Roselle Abraham, and HagitShatkay et al. “Utilizing ECG-based Heartbeat Classification for Hypertrophic Cardiomyopathy Identification”, DOI 10.1109/TNB.2015.2426213, IEEE Transactions on Nano Bioscience TNB-00035-2015.
[53] RenuChauhan, Pinki Bajaj, KavitaChoudhary and YogitaGigras et al. “Framework to Predict Health Diseases Using Attribute Selection Mechanism”, 2015 2nd International Conference on Computing for Sustainable Global Development (INDIA Com).
[54] Rahul Kumar, Ankit Agrawal, and Alok Kumar. (2023). Heart disease prediction using machine learning algorithms: A review. International Journal of Engineering and Technology, 10(5), 2877-2883. doi:10.14419/ijet.v10i5.5133

[55] R. Vijaya Saraswathi, Kovid Gajavelly, A. Kousar Nikath, R. Vasavi,and Rakshith Reddy Anumasula Heart Disease Prediction Using Decision Tree and SVM Feb(2022)
[56] Ravindhar NV, Anand, HariharanShanmugasundaram, Ragavendran, Godfrey Winster. Intelligent Diagnosis of Cardiac Disease Prediction Using Machine Learning. Volume-8 Issue-11, September 2019, ISSN: 2278-3075 (Online). Page No: 1417-1421. DOI: 10.35940/ijitee.J9765.0981119 https://archive.ics.uci.edu/ml/datasets/Heart+Disease
[57] Shubham Sharma, Rahul Gupta, and Amit Kumar. (2022). Heart disease prediction using machine learning algorithms: A comparative study. Journal of Medical Imaging and Health Informatics, 12(2), 371-377. doi:10.1166/jmihi.2022.2876
[58] Shah, D., Patel, S. &Bharti, S.K. Heart Disease Prediction using Machine Learning Techniques.SN COMPUT. SCI. 1, 345 (2020).https://doi.org/10.1007/s42979-020-00365
[59] Shan Xu, Tiangang Zhu, Zhen Zang, Daoxian Wang, Junfeng Hu and XiaohuiDuan et al. “Cardiovascular Risk Prediction Method Based on CFS Subset Evaluation and Random Forest Classification Framework”, 2017 IEEE 2nd International Conference on Big Data Analysis.
[60] SeyedaminPouriyeh, Sara Vahid, Giovanna Sannino, Giuseppe De Pietro, Hamid Arabnia, Juan Gutierrez et al. “A Comprehensive Investigation and Comparison of Machine Learning Techniques in the Domain of Heart Disease”, 22nd IEEE Symposium on Computers and Communication (ISCC 2017): Workshops - ICTS4eHealth 2017.
[61] Shakti Chourasiya and Suvrat Jain, "A Study Review on Supervised Machine Learning Algorithms," (SSRGIJCSE), vol. 6, no. 8, 2019.
[62] SeyedaminPouriyeh, Sara Vahid, Giovanna Sannino, Giuseppe De Pietro, Hamid Arabnia, Juan Gutierrez, et al. “A Comprehensive Investigation and Comparison of Machine Learning Techniques in the Domain of
[63] Simge EKIZ and PakizeErdogmus et al. “Comparative Study of heart Disease Classification”, 978-1-5386-0440-3/17/$31.00 ©2017 IEEE.
[64] S.Rajathi and Dr.G.Radhamani et al. “Prediction and Analysis of Rheumatic Heart Disease using kNN Classification with ACO “, 2016.
[65] S. Hochreiter and J. UrgenSchmidhuber, “Long Short-Term Memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.
[66] S. Palaniappan and R. Awang, “Intelligent heart disease prediction system using data mining techniques,” 2008 IEEE/ACS Int. Conf. Comput. Syst. Appl., pp. 108–115, 2008.
[67] Subbalakshmi, G., Ramesh, K., &ChinnaRao, M. (May 2011). Decision Support in Heart Disease Prediction System using Naive Bayes. Unpublished research paper. Kakinada Institute of Engineering & Technology, Yanam Road, Korangi-533461, E.G.Dist., A.P., India.
[68] T. A. Lasko, J. C. Denny, and M. A. Levy, “Computational Phenotype Discovery Using Unsupervised Feature Learning over Noisy, Sparse, and Irregular Clinical Data,” PLoS One, vol. 8, no. 6, 2013.
[69] Thenmozhi, K., &Deepika, P. (2014). Heart Disease Prediction Using Classification with Different Decision Tree Techniques. International Journal of Engineering Research and General Science, 2(6), 6-10. Retrieved from http://www.ijergs.org/volume2_issue6/IJERGS020607.pdf.
[70] V. Kirubha and S. M. Priya, “Survey on Data Mining Algorithms in Disease Prediction,” vol. 38, no. 3, pp. 124–128, 2016
[71] V.Manikandan and S.Latha, " Predicting the Analysis of Heart Disease Symptoms Using Medical Data Mining Methods" , International Journal of Advanced Computer Theory and Engineering, Vol. 2, Issue. 2, 2013
[72] Venkat, V., Abdelhalim, H., DeGroat, W., Zeeshan, S., & Ahmed, Z. (2023). Investigating genes associated with heart failure, atrial fibrillation, and other cardiovascular diseases, and predicting disease using machine learning techniques for translational research and precision medicine. Genomics, 115(2), 110584.
[73] YumingHua, JunhaiGuo, and Hua Zhao, “Deep Belief Networks and deep learning,” Proc. 2015 Int. Conf. Intell. Comput.Internet Things, pp. 1–4, 2015.

Authors Profile

Mrs. Parvati Kanaki obtained Master in Electronics and Communication Engineering from Visvesvaraya Technological University, Belagavi, India. She has made significant contribution in carrying out several research papers in Biomedical Image Processing, Standard clinical ECG, ML Algorithms, Logistic Regression, Random Forest, SVM, Genetic algorithm, Networking.

Dr. Gyanappa A. Walikar obtained both Master and Ph.D. in Computer Science & Engineering from Visvesvaraya Technological University, Belagavi, India. He has made significant contribution in carrying out several research papers and projects on Design and Development of Hybrid Multicast Routing Schemes in MANET. Some of the Journals where his research articles published are Elsevier, Inderscience, and IEEE Conferences. He is the recipient of Reviewer Award from various reputed Journals and Conferences like, Elsevier, Springer, Wiley, Open Science Journal, and IEEE Conferences. Besides, he is a Member of ISTE, CSI, and INAAR. He chaired several sessions at National and International Conferences. He worked as a Technical Committee Member, Advisory Member at International Conferences. He has also been working as an Editorial Member of reputed & scholarly journals.