HEALTH PULSE PROGNOSIS: PREDICTING HEART DISEASE THROUGH INTELLIGENT DATA MINING STRATEGIES

Abstract

In the modern day, fatalities that are caused by heart disease have become a significant problem. It is estimated that nearly one person passes away due to heart disease every minute. In the majority of situations, the diagnosis of cardiac disease is dependent on a complicated mix of clinical and pathological data. This complexity results in exorbitant medical expenses, which in turn affects the quality of medical care that is provided. The objective of the research is to evaluate medical information in order to make a prediction regarding heart disease. In order to discover information contained within a dataset pertaining to heart illness, which was retrieved from the UCI Machine Learning repository, the suggested technique makes use of data mining classification algorithms such as KNN, SVM, and Decision Tree. In all, there are 282 observations and 75 characteristics that make up the input dataset. The feature selection of the input dataset is accomplished through the utilisation auto-correlation in order to locate the most optimised collection of features and enhance the overall performance of the classifiers. When it comes to accuracy, the results are achieved through the use of classification techniques, both with and without the utilization of feature selection. The final step is to make a prediction about the class labels by employing the Majority Voting method.

Keywords: Heart disease, auto correlation, data mining, disease diagonsis, feature selection, classification and prediction;

Authors

D. Prabhu
Assistant Professor
Department of Computer Science and Engineering
School of Computing
Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology
Deemed to Be University
Avadi, Chennai, India.
prabhu.klnce@gmail.com

T. Anand
Assistant Professor
Department of Computer Science and Engineering
School of Computing
Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology
Deemed to Be University
Avadi, Chennai, India.
t.anandme@gmail.com

Dr. D. Poornima
Assistant Professor
Department of Computer Science and Engineering
Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu India.
poornima.d.cse@sathyabama.ac.in

Ms. M. P. Bobby
Assistant Professor
Department of Computer Science and Engineering
Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu India.
bobby.m.p.cse@sathyabama.ac.in

I. INTRODUCTION

A significant portion of our study is concentrated on a variety of data mining approaches that are helpful in the prediction of cardiac disease. If the heart is not functioning properly, it will cause other areas of the human body, such as the kidneys, the brain, and other organs, to experience discomfort. A type of sickness that interferes with the normal functioning of the heart is referred to as heart disease. In our day and age, the most common cause of mortality is diagnosed as heart disease. According to projections made by the World Health Organisation (WHO), cardiac illnesses should be responsible for the deaths of 12 million people annually.

The symptoms of cardiac disease might vary from person to person, depending on the individual's biology, and these symptoms can also change. The majority of the time, however, they consist of chest discomfort, pain in the arms and shoulders, pain in the back, pain in the jaw, pain in the neck, stomach issues, and problems with breathing. Heart failure, stroke, and coronary artery disease are just few of the cardiac conditions that may be found among the many different types of heart disorders. The two primary causes of heart disease director are a healthy lifestyle (principal prevention) and prompt analysis (inferior prevention). Both of these factors are important in primary prevention. The practice of doing routine checkups, also known as inferior preventive, plays a significant role in the diagnosis and early prevention of heart disease related complications. A number of diagnostic procedures, including angiography, chest X-rays, echocardiography, and exercise tolerance testing, provide evidence in support of this critical concern.

With that being said, these examinations are not only costly but also need the availability of precise medical equipment. Professionals in the field of heart care establish a comprehensive and extensive database of patients and keep it. In addition to this, it offers a fantastic opportunity to extract valuable information from datasets of this kind. It is important that the data that is utilised in the dataset be full and consistent. Therefore, make sure that the data set does not contain any data that is inconsistent, noisy, or incomplete. In order to assist experts or physicians in identifying cardiac illness, researchers make use of a variety of data mining tools that are readily available. Decision trees, k-nearest neighbours (KNN), and support vector machines (SVM) are examples of examples of techniques that are often utilised. The process of detecting patterns in huge data sets is referred to as data mining.

This process involves the use of techniques that lie at the confluence of machine learning, statistics, and database management systems. Generalisation, characterisation, classification, clustering, association mining, pattern matching, data visualisation, and meta-rule guided mining are some of the approaches that have been developed in recent years for the purpose of data mining. The findings of this study provide a strategy for forecasting the illness of the heart. On the basis of the medical data, a variety of categorization strategies are utilised in order to make a prediction regarding heart disease.

This work is organised in the following manner for the remaining parts: A overview of research that are pertinent is presented in Chapter 2. Chapter 3 provides an explanation of the technique that is being offered. There is a description of the implementation and the work that was done in Chapter 4.Chapter 5 contains the narration of the conclusion.

II. LITERARTURE SURVEY

Over the course of the past several years, numerous hospitals have conducted a multitude of research on the anticipation of a variety of diseases. A variety of data mining strategies were utilised for the purpose of diagnosis, with variable degrees of success. The section that follows contains a concise literature overview of the various methods of data mining and assessment that are utilised in the healthcare industry. In order to accomplish the goal of this project, which is to create an automated medical heart disease prediction system, the obtained database will be carefully analysed. A medical examination of cardiac disorders can be carried out in two stages, according to the patient's condition. There are a number of different algorithms that have been developed for the pre-processing stage, which is the first one. The pre-processing step's purpose is to choose just the necessary features from the entire collection of database. Both the filter approach and the wrapper method are the two feature selection strategies that are utilised the most often. The methods for constructing the prediction model by using classifier are included in the second phase of the analysis section of the process. In general, the classifier is divided into two categories: those that are based on machine learning and those that are based on fuzzy systems according to the technology that is utilised.

A number of different machine learning strategies for the prediction of cardiac illnesses have been presented, all of which are based on a single classifier. For the purpose of prediction, single classifiers such as SVM, Decision Tree, Neural Network, Naives Bayes, and Regression are utilised. For the purpose of illness prediction, a machine learning model that is based on supervised learning techniques [7] has been presented. A number of different algorithms, including Naïve Bayes, decision tree, K-nearest neighbour, and random forest algorithms, have been utilised in the development of the prediction model.

For the purpose of making more accurate predictions regarding illnesses, the ensemble classifier is the one that will be employed. This classifier will consist of more than one classifier. For the purpose of identifying earlier stages of cardiac disease, the Ensemble learning approaches using Particle Swarm Optimisation (PSO) [1] have been proposed. In this article, a strategy for selecting feature subsets that is based on both filter and wrapper approaches is described and then implemented using the PSO technique [2].

In order to make accurate predictions regarding cardiac disorders, ensemble approaches like as Bagging, Random Forest, and AdaBoost are utilised. When it comes to fitting a model to bins, the signal-to-noise ratio is an extremely important factor. It is possible for even a slight variation in the data of the model to lessen the impact of fitting the model. In order to increase the signal-to-noise ratio and also to categorise the subset characteristics that are required for accurate prediction from a large quantity of data, a data reduction process that is based on discretization [3] has been developed. Once this is done, the ensemble classifier, which consists of Bagging, Random Forest, and Voting ensemble, is utilised in order to construct the most accurate prediction model possible. Eighty percent of the data that was gathered is used as a training dataset, and twenty percent of the data is regarded to be a test dataset for the purpose of evaluating the prediction model.

It is recommended that an ensemble classifier be developed using a machine learning technique that is based on the Adaptive Boosting method [4]. The method is applied to four distinct datasets, and the results demonstrate that the ensemble classifier that is based on adaptive boosting algorithms has reached higher levels of dependability and delivers levels of accuracy that are superior to those produced by earlier studies. One of the models that is utilised in clinical settings all around the world by professionals in the field of cardiac illnesses is the prediction model that is utilised in this technique [5-7].

Improvements in the performance of the prediction mechanism may be achieved by the extraction of features through the use of LDA and PCA techniques [8]. In the early days of algorithm development, the ensemble classifier algorithms that were utilised the most frequently were bootstrapping aggregation [9], boosting [10], and random decision forests [11]. When these ensemble classifiers are utilised, there has been an improvement in the performance of the classification efficiency. It has been suggested that in order to accomplish ensemble learning, a system that integrates a large number of classifiers or partitions based on voting techniques should be utilized [12].

Voting is a term that refers to the process of picking the best answer from among the available options or developing views that result in a consensus among the participants. Through the process of voting, every voter has the opportunity to express their preference for one or more candidates [13]. The candidate who is found to be the most desirable is chosen to be the best classifier. Various methodologies, including parametric, nonparametric, heuristics, logical, and probabilistic approaches, can be utilised in the construction of voting systems. Voting systems may be used for a variety of purposes, including selecting the most effective classifier [14]. One innovative method involves combining a voting system with a number of different machine learning algorithms and then choosing the most effective one. A number of different classifiers, including decision trees, k-nearest neighbours, and multilayer perception, are selected by the machine learning algorithms, and they are all trained using the same data set [15]. In a different manner, the voting system may be utilised with a single classifier, but it can be trained using a separate data set.

It is decided to conduct a survey in order to investigate the impact that voting methods [16] have on categorization tasks. To investigate the influence of voting approaches on the performance of the k-Nearest Neighbour and Naïve Bayes algorithms, a variety of datasets are utilised for the purpose of conducting statistical analysis. It is decided to do a comparative analysis of the ensemble decision tree approaches [17]. In order to carry out the studies, ensemble decision tree methods, such as bootstrap aggregating and server randomization, were utilised. After careful analysis, it was shown that random forests, boosting, and randomised tree performed better than bagging. There are a variety of basic classifiers that may be utilised with the ensemble learning approach. Ensembles may be built by first splitting the dataset into subgroups and then mixing those subsets together [18, 19].

The formation of the additional datasets was accomplished by the use of random permutation [20] and dividing of the first dataset. The resultant ensemble achieved a high level of generalisation as a consequence of the overall compatibility of the various models being greatly improved. [21, 22]. The use of ensemble learning has been utilised in the prediction of a number of medicinal treatments. An ensemble model for the prediction of cardiac illness is suggested in this study. The model makes use of machine learning classifier methods, namely support vector machines (SVM), decision trees, and random forests [23]. The research efforts are being done with the intention of selecting the relevant characteristics in order to improve the accuracy of the prediction while simultaneously reducing the amount of computer resources required. This is accomplished by the use of feature selection methods, auto correlation, which are mostly employed for the purpose of feature reduction. The output class is predicted by the voting classifier based on the one that received the biggest majority of votes. The predictions of each base classifier are then sent into the voting classifier. Creating a single model by training several base classifiers and predicting the output based on the cumulative majority of votes for each output class is the primary objective of this. This involves training multiple base classifiers.

III. PROPOSED METHODOLOGIES

The heart disease diagnosis system was built using voting ensemble classifiers and auto correlation feature selection methods, as described in the section that follows. The UCI Machine Learning Repository, specifically the heart disease dataset, is used to compile the 82 observations and 75 variables that make up the Cleveland dataset. With 75 characteristics in the original dataset, 10 features were identified and selected using an auto correlation feature selection approach. The modified dataset is then fed into classification algorithms like KNN, Decision Tree, and SVM. In order to examine the results, vote ensemble classification algorithms are used. The heart disease prediction system's architecture is shown in Figure 1.

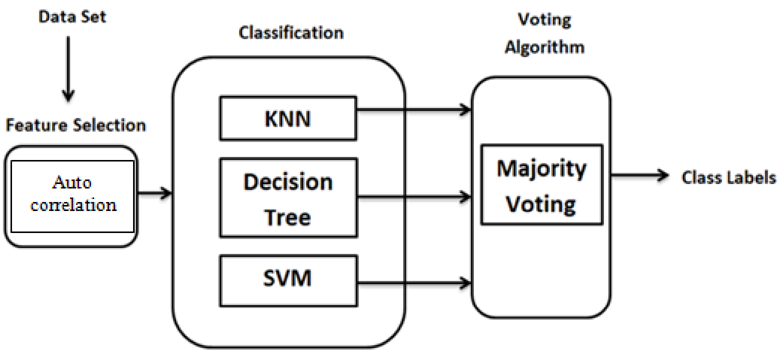

Figure 1: Architecture of Heart Disease Diagnosis Systems

The Heart disease system prediction system consists of flowing modules:

- Dataset Collection
- Feature Selection
- Classification
- Voting Classification
- Final Prediction

1. Dataset Collection

The heart disease dataset that is provided in the UCI Machine Learning repository is the source of the 282 observations and 75 characteristics that make up the input dataset.The dataset that was collected includes information such as the patient's gender, age, kind of chest discomfort, resting blood pressure, resting electrocardiogram findings, and the names of the variables that are associated with each individual patient.The dataset that was gathered is now accessible in a processed form that contains data that is both comprehensive and consistent. This sample dataset's contents are presented in Table 1.

Table 1: Dataset Attribute Information

S.No.	Attributes	Description	Values
1	Age	Patient's age in years	Continuous Value
2	Sex	Sex of Patient	1=Male 0=Female
3	Trestbps	Resting blood pressure	Continuous value in mm/Hg
4	Chol	Serum cholesterol in mg/dl	Continuous value in mg/dl
5	Fbs	Fasting blood sugar	1>=120 mg/dl 0<=120 mg/dl
6	Thalach	Maximum heart rate achieved	Continuous value

2. Feature Selection

The process of picking the subset of the features that make up the set of features that are the most relevant is referred to as feature selection. In this work, we have used auto correation as feature selection. Autocorrelation refers to the property of data where the correlation between values of the same variables is determined by their relationship to each other. This contradicts the notion of instance independence, which is a fundamental principle in most traditional models. It typically occurs in datasets where the data is sourced from the same origin, rather than being randomly picked.

The original dataset is subjected to the Correlation based feature selection (CFS) process to identify class characteristics that are relevant.Collinear characteristics are characterised by a strong correlation between them. These factors in machine learning result in a reduction in the overall performance of the model on the test set due to increased variability and lower ability to comprehend the model. First-order serial correlation, which can be either positive or negative, is the most prevalent kind of autocorrelation. Positive serial correlation refers to a situation in which a positive deviation in one period persists as a positive deviation in the subsequent period.Negative serial correlation refers to the phenomenon where a negative deviation in one time period persists and leads to a negative deviation in the subsequent time period.

3. SVM Classifer

In the field of Supervised Learning, one of the most often used methods is known as the Support Vector Machine (SVM). The classification and regression tasks are two of the most prominent applications for this technique. Nevertheless, its principal use is in the identification and resolution of categorization issues within the realm of machine learning.

An ideal line or decision boundary that successfully divides an n-dimensional space into different classes is the goal of the Support Vector Machine (SVM) approach. This technique was developed with the intention of achieving this aim. Because of this, we are able to properly categorise future data points into the categories that are most suited for them. A hyperplane is the name given to the decision boundary that ensures the data points are separated in the most effective manner. The most important points and vectors that contribute to the construction of the hyperplane are chosen by the support vector machine (SVM). Support Vector Machine is a method that incorporates extreme conditions that are known as support vectors. The word "Support Vector Machine" alludes to this method. The presence of two separate categories that are divided from one another by a decision boundary or an ideal hyperplane is depicted in Figure 2.

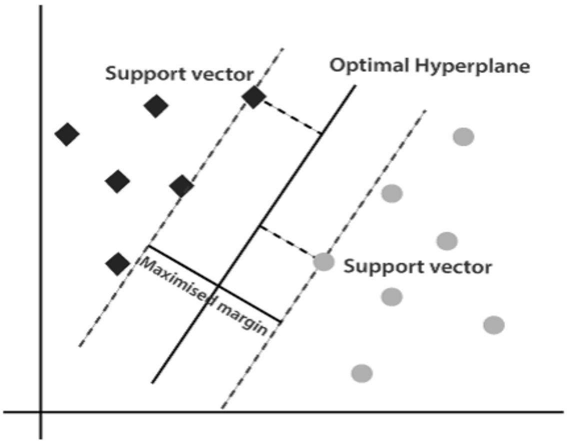

Figure 2: SVM and Hyperplane

4. kNN Classifier

The K nearest neighbour algorithm is a method that saves all existing instances and categorises new cases by using a similarity metric, such as distance functions.The K Nearest Neighbour (KNN) algorithm is a commonly employed classification method.There is either no clear training period or it is quite limited. Additionally, this implies that the training period is rather rapid. KNN's lack of generalisation is due to its retention of all the training data. Precisely, the testing phase requires the use of all (or the most of) the training data. The KNN algorithm relies on the similarity of features. The degree of similarity between out-of-sample attributes and our training set impacts the classification of a particular data point. Figure 3 illustrates the process of categorising a new data point using the KNN algorithm.

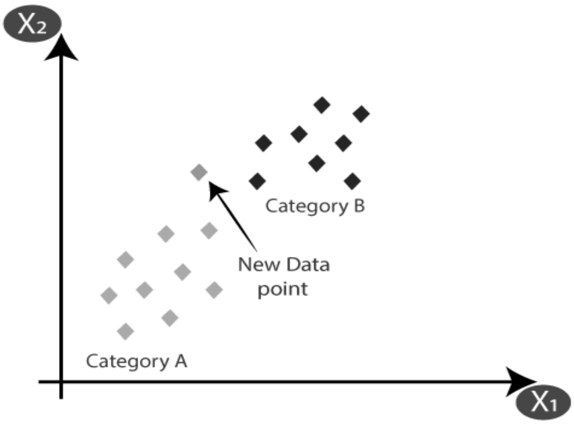

Figure 3: Knn Classifier Separation

5. Decision Tree Classifier

An example of a decision support tool is a decision tree, which is a tree-like graph or model of alternatives and the probable consequences of those decisions. These outcomes may include the results of random occurrences, the costs of resources, and the usefulness of those resources. Decision trees are a form of decision support tool. A method that only comprises statements that control the conditions is one way to display an algorithm. This is one technique. A decision tree is a structure that is similar to a flowchart in that each internal node represents a "test" on an attribute, each branch shows the outcome of the test, and each leaf node represents a different class label. In other words, a decision tree is a structure that is an example of a flowchart. The paths that extend from the root to the leaf are the pathways that reflect the classes. On display in Figure 4 is a representation and illustration of the creation of the Decision Tree.

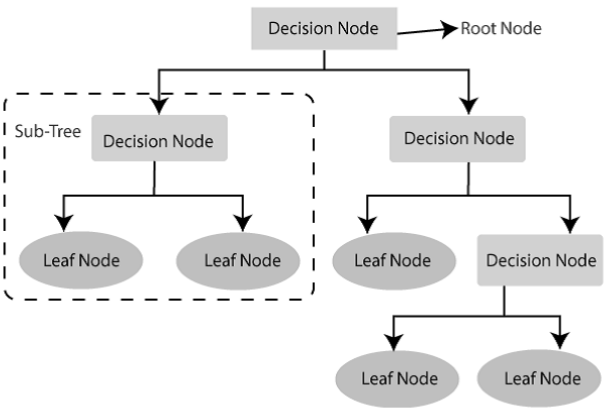

Figure 4: Decision Tree Classifier Separation

6. Voting Classifier

The techniques of voting are based on a weighted procedure that combines the predictions that are offered by the classification models that have been separately calibrated on a number of different analytical sources that are available. The class that has the highest number of predictions across all of the classifiers is the one that is selected to be the ensemble's prediction through the procedure known as majority voting. The method of voting from the majority is depicted in Figure 5.

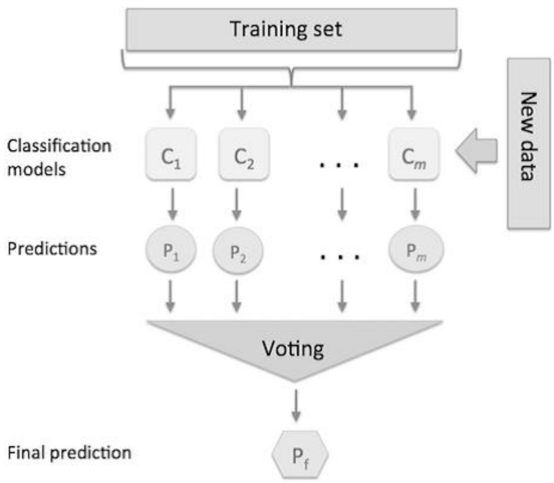

Figure 5: Ensemble Voting Classifier

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

An autocorrelation analysis yielded a total of nine characteristics as a result. Cholesterol, the number of years spent smoking (years), resting electrocardiogram (restecg), the duration of the exercise test in minutes (thaldur), the maximum heart rate achieved (thalach), the number of major vessels coloured by flouroscopy (Ca), rest wall motion abnormality (restwm), along with thal, thalpul, and laddist are the ten features that were extracted. A representation of the code for auto correlation may be found in Figure 6.

```
# Select K best features
k_best_features = SelectKBest(mutual_info_classif, k=10).fit(X_train, y_train)
print('Selected top 10 features: {}'.format(X_train.columns[k_best_features.get_support()]))

Selected top 10 features: Index(['Chol', 'years', 'Restecg', 'Thaldur', 'Thalach', 'Ca', 'Restwm',
       'Thal', 'thalpul', 'Laddist'],
      dtype='object')
```

Figure 6: Feature Selection using Auto Correlations

The KNN classifier searches for the classes of the K closest neighbours of a particular data point, and then assigns a class to the data point based on the class that is the majority of the neighbours. As shown in Figure 7, an accuracy of 83.55% was achieved by the utilisation of KNN data mining classification using 10 features.

```
k=KNeighborsClassifier(algorithm='auto', n_neighbors= 21)
score_k=CrossVal(Xtrain,Ytrain,k)
print("Accuracy is : ",score_k)
k.fit(Xtrain,Ytrain)
plotting(Ytest,k.predict_proba(Xtest))


fig=plt.figure()
sns.heatmap(confusion_matrix(Ytest,k.predict(Xtest)), annot= True, cmap='Reds')
```

```
Accuracy is :  0.8355555555555554
```

Figure 7: Heart Disease Prediction using k-NN Classifiers

The Decision Tree classifier is responsible for the creation of a decision tree, which serves as the basis for the assignment of class values to each different data point. As shown in Figure 8, with 10 characteristics the accuracy that can be achieved via the utilisation of decision tree data mining categorization accounts for 78.66%.

```
dtc=DecisionTreeClassifier()
score_dtc=CrossVal(Xtrain,Ytrain,dtc)
print("Accuracy is : ",score_dtc)
dtc.fit(Xtrain,Ytrain)
plotting(Ytest,dtc.predict_proba(Xtest))

fig=plt.figure()
sns.heatmap(confusion_matrix(Ytest,dtc.predict(Xtest)), annot= True, cmap='Blues')
```

```
Accuracy is :  0.7866666666666667
```

Figure 8: Heart Disease Prediction using Decision Tree classifiers

Through the process of adjusting the distance between the data points and the hyperplane, the support vector machine (SVM) classifier strives to create a hyperplane that is capable of separating the classes to the greatest extent feasible. The hyperplane is determined by a number of different kernels, of which there are numerous. For the purpose of determining the hyperplane, this project makes use of the linear kernel. As shown in Figure 9, 85.33% is the accuracy that can be achieved with the utilisation of SVM data mining classification using 10 features.

```
svc=SVC(C=0.2,probability=True,kernel='rbf',gamma=0.1)
score_svc=CrossVal(Xtrain,Ytrain,svc)
print("Accuracy is : ",score_svc)
svc.fit(Xtrain,Ytrain)
plotting(Ytest,svc.predict_proba(Xtest))

fig=plt.figure()
sns.heatmap(confusion_matrix(Ytest,svc.predict(Xtest)), annot= True, cmap='Greys')
```

```
Accuracy is :  0.8533333333333333
```

Figure 9: Heart Disease Prediction using SVM Classifiers

```
from mlxtend.classifier import EnsembleVoteClassifier
from sklearn import model_selection
eclf = EnsembleVoteClassifier(clfs=[k, dtc, svc])

labels = ['Knn', 'DTC', 'SVC', 'Ensemble']
for clf, label in zip([k, dtc, svc, eclf], labels):

    scores = model_selection.cross_val_score(clf, Xtrain, Ytrain,
                                             cv=5,
                                             scoring='accuracy')
    print("Accuracy: %0.2f (+/- %0.2f) [%s]"
          % (scores.mean(), scores.std(), label))
```

```
Accuracy: 0.84 (+/- 0.04) [Knn]
Accuracy: 0.76 (+/- 0.07) [DTC]
Accuracy: 0.85 (+/- 0.04) [SVC]
Accuracy: 0.86 (+/- 0.02) [Ensemble]
```

Figure 10: Heart Disease Prediction using Ensemble Voting Classifiers

In order to the Heart disease predctions, we have combined the three classifiers namely support vector machne, decision tree and k-Nearest Neighbors and calculate the voting ensemble stategies. As Shown in Figure 10, the voing classifiers have superier result rather than the other three classifiers. The consuion matrix of three traditional classifier as well as ensmble voring classifier are shown in Figure 11.

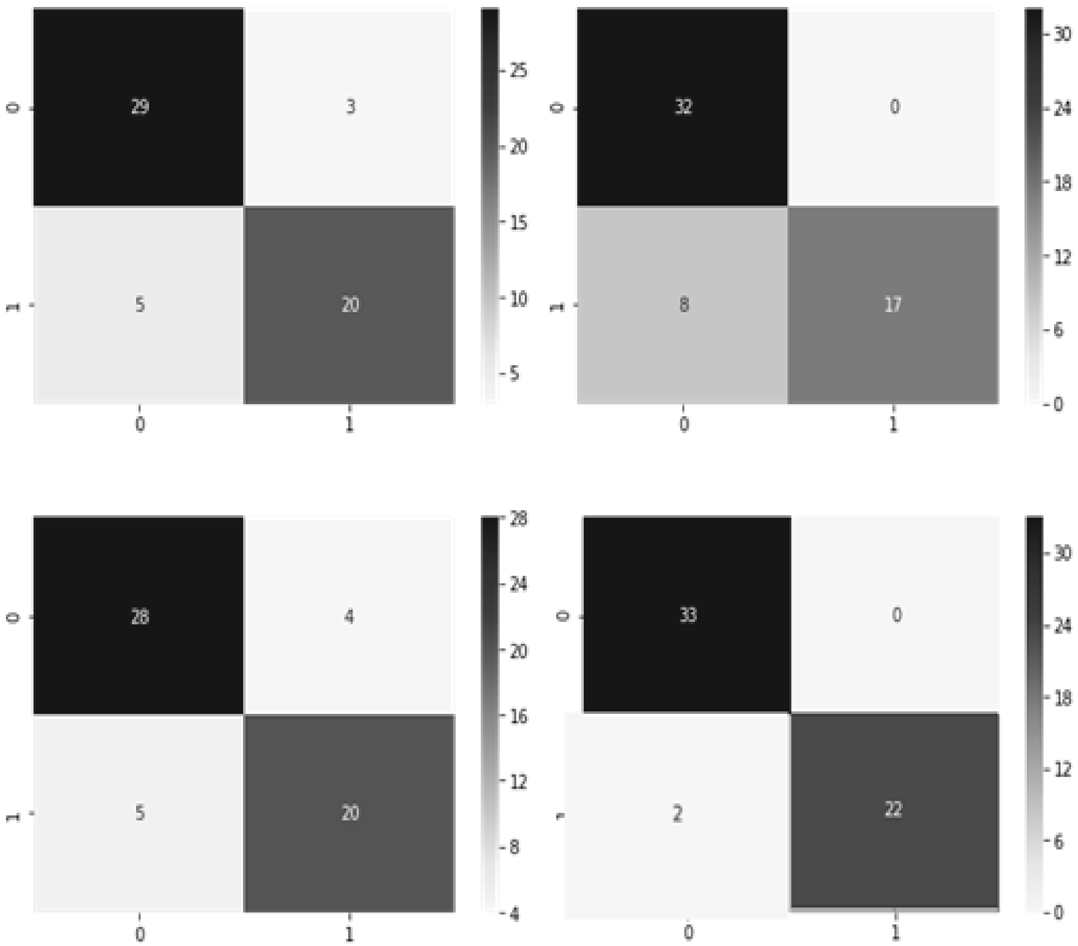

Figure 11: Confuson Matrix of all the Classifiers

Table 2: Comparative Analysis of Traditional and Ensemble Classifier

S. No.	Classifier	Accuracy	Precision	Recall	F1-Score
1.	SVM	85.3%	88.5%	86.6%	87.1%
2.	DT	76.4%	75.4%	73.7%	76.2%
3.	KNN	83.5%	92.9%	93%	85%
4.	Voting Classifier	87.2%	94.5%	94.7%	91.3%

V. CONCLUSIONS

The accuracy of heart disease prediction using an ensemble of classifiers is examined in this study. For training and testing, we used the Cleveland heart dataset that is part of the UCI machine learning library. In order to classify databases pertaining to cardiac diseases, we have investigated many classification techniques. Additionally, we have examined several categorization methods and the results they provide. Depending on the execution tools, this analysis reveals distinct technologies utilised in different publications with varied counts of characteristics and varying degrees of accuracy. Developing new permutations of data mining techniques can further improve the structure's accuracy. An accuracy of 83.5% was achieved when KNN was employed. A 78.6% success rate was achieved when a decision tree was employed.

With SVM, the accuracy reached 85.3%. When looking at the results side by side, it was clear that majority voting significantly improves accuracy. Using feature selection techniques substantially improved the performance. The accuracy of the ensemble algorithms was enhanced by the feature selection strategies. We discovered and put into action the Heart Disease Prediction System that makes use of a machine learning algorithm. The proposed system makes use of a multiple feature assessment technique and majority voting procedure for class label prediction, which are both efficient and accurate, due to the current technological developments that have led to a great deal of evolution in machine learning algorithms. Based on user input, the system also provides nearby trustworthy results. A decrease in the mortality rate from cardiovascular illnesses is possible if more people use the system and are therefore aware of their present heart health.

REFERENCES

[1] Y. Indu, D. Sunanda and M. Jabbar, Prediction of heart disease using ensemble learning and Particle Swarm Optimization, p. 691-698, (2017).

[2] L. Ladha and T. Deepa, Feature Selection Methods and Algorithms, International Journal of CSE, p.1787-1797, (2011).

[3] J. Emakhu, S. Shrestha and S. Arslanturk, Prediction System for Heart Disease Based on Ensemble Classifiers, Proceedings of the 5thInternational Conference on Industrial Engineering and Operations Management Detroit, Michigan, USA, August p. 10 - 14, (2020).

[4] H. Kathleen, H. Julia, and J. George, Diagnosing Coronary Heart Disease Using Ensemble Machine Learning Diagnosing Coronary Heart Disease Using Ensemble Machine Learning, International Journal of Advanced Computer Science and Applications, p. 1-12, (2016).

[5] K. Mohamad Almustafa, Prediction of heart disease and classifiers sensitivity analysis, BMC Bioinformatics, p. 1-5, (2020).

[6] X. Gao, A. Amin Ali, H. Shaban Hassan and E.M. Anwar, Improving the Accuracy for Analyzing Heart Diseases Prediction Based on the Ensemble Method, p. 1-12, (2021).

[7] D. Shah, S. Patel, and S.K. Bharti, Heart Disease Prediction using Machine Learning Techniques, Soft Neuro Computing, p. 15-21, (2020).
[8] I.D. Mienye, Y. Sun, Z. Wang, An improved ensemble learning approach for the prediction of heart disease risk, Informatics in Medicine Unlocked, p. 312-322, (2020).
[9] T.K. Ho, The random subspace method for constructing decision forests, IEEE Transaction on Pattern Analysis Mach Intelligence, p. 832-844, (1998).
[10] L. Breiman, Bagging predictors, Mach Learning, p. 123-140, (1996).
[11] R.E. Schapire and Y. Singer, Improved boosting algorithms using confidence-rated predictions, Mach Learn, p. 297-336, (2011).
[12] F. Leon, S.A. Floria, C. Badica, Evaluating the effect of voting methods on ensemble-based classification, IEEE International Conference on Innovations in intelligent Systems and Applications (INISTA), p. 1-6, (2017).
[13] R.E. Banfield, L.O. Hall, K.W. Bowyer and W.P. Kegelmeyer, A comparison of decision tree ensemble creation techniques, IEEE Trans Pattern Anal Mach Intell, p. 173-180, (2017).
[14] D. Ruta, B. Gabrys, C. Lemke, A generic multilevel architecture for time series prediction, IEEE Transaction on Knowledge Data Engineering, p. 350-359, (2011).
[15] B. Zhang, Ensemble learners of multiple deep CNNs for pulmonary nodules classification using CT images, IEEE Access, p. 110358-110371, (2019).
[16] L. Han, S. Luo, J. Yu, L. Pan, S. Chen, Rule extraction from support vector machines using ensemble learning approach: an application for diagnosis of diabetes, IEEE Journal of Biomedical and Health Informatics, p. 728-734, (2015).
[17] J. Emakhu, S. Shrestha and S. Arslanturk, Prediction System for Heart Disease Based on Ensemble Classifiers, Proceedings of the 5thInternational Conference on Industrial Engineering and Operations Management Detroit, Michigan, USA, August p. 10 - 14, (2020).
[18] H. Kathleen, H. Julia, and J. George, Diagnosing Coronary Heart Disease Using Ensemble Machine Learning Diagnosing Coronary Heart Disease Using Ensemble Machine Learning, International Journal of Advanced Computer Science and Applications, p. 1-12, (2016).
[19] K. Mohamad Almustafa, Prediction of heart disease and classifiers sensitivity analysis, BMC Bioinformatics, p. 1-5, (2020).
[20] X. Gao, A. Amin Ali, H. Shaban Hassan and E.M. Anwar, Improving the Accuracy for Analyzing Heart Diseases Prediction Based on the Ensemble Method, p. 1-12, (2021).
[21] Pati, A., Parhi, M., & Pattanayak, B. K. (2022). IHDPM: An integrated heart disease prediction model for heart disease prediction. International Journal of Medical Engineering and Informatics, 14(6), PP. 1-12.
[22] Shuvo, S. (2023). An early detection of heart disease using machine Learning (recurrent neural network). https://doi.org/10.14293/s2199-1006.1.sor-.ppgv4id.v1
[23] VARKALA, K. (2022). Heart disease prediction system using Convolutional neural networks. https://doi.org/10.21203/rs.3.rs-2009078/v2

