ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Issue 12, December 2021

Heart Disease Prediction Using Hybrid KNN and Random Forest Algorithm

P. Bavithra[1]

[1]Research Scholar in Computer Science
Department of Computer Science, Jamal Mohamed College (Autonomous)
(Affiliated to Bharathidasan University)
Tiruchirappalli 620 020, Tamil Nadu, India
[1]bavithrapandiyan@gmail.com

Abstract: The heart is vital organ of human body part which pumps blood throughout the body. If blood circulation in the body is ineffective the organs like brain suffer and if heart discontinues working altogether, death happens within minutes. Life is completely dependent on proficient working of the heart. The term Heart disease refers to disease of heart & blood vessel system in it. Data mining it is a process of examining data from dissimilar perspective and gathering the knowledge from it. The discovered knowledge can be used for dissimilar applications for example healthcare industry. These days healthcare industry produces huge amount of data about patients, disease diagnosis etc. Data mining gives a set of techniques to determine hidden patterns from data. Heart disease (HD) is a main cause of morbidity and mortality in the modern society. Medical diagnosis is a significant but difficult task that should be performed precisely and efficiently and its automation would be very helpful. In this research work, we develop a heart disease prediction system that can assist medical professionals in calculating heart disease status based on the clinical data of patients. Processing two algorithms within a single workflow is called as hybrid approach (combined method), in this research a hybrid of K-Nearest Neighbor (KNN) and Random Forest Classifier is proposed. This proposed hybrid approach provides optimum solution for predicting the Heart diseases using two datasets (Cleveland and Framingham) that are collected from UCI and Kaggle Repository. The experimental result of the proposed work is compared with the result of Decision Tree, KNN, Naive Bayes, Logistic Regression, Random Forest and SVM classification algorithms. Prediction of cardiovascular disease is a vital challenge in the area of clinical data analysis. Our proposed hybrid approach produces an enhanced performance with high level of accuracy.

Keywords: Data Mining, Hybrid, Classification, Machine Learning Techniques, Heart Disease Prediction.

I. INTRODUCTION

Health services are an obvious task to be done in human life. Human being's concern business has become a notable field in the wide territory of medical science. The health services industry contains a huge measure of information and hidden data. Compelling choices are made with this unknown data by applying data mining methods. Numerous tests are done in the identification of cardiovascular illnesses in the patient anyway with data mining these tests could be reduced. However, there is a lack of analysis device to furnish successful test results with the concealed data, so a framework is created using data mining techniques for classifying the information and to identify the heart illnesses. Data mining goes about as an answer to some health problems.

In general, Data mining is known as "an abstraction process of implicit, formerly unidentified and possibly beneficial information from the data stored during a database" or as "a process of selection, exploration, and modelling of huge quantities of data to urge regularities or relations that are primarily unknown with the aim of obtaining clear and useful outcomes for the owner of the database" [1]. Data mining holds huge potential in healthcare industry to allow health systems to systematically use data and analytics to recognize inefficiencies and best practices that increase care and reduce costs.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Issue 12, December 2021

The risk of heart disease is much high all over the world and is a main cause of death. Predicting the disease is hard as it needs much knowledge and experience. It is a very tough process even if the disease is predicted as it takes quiet a long time to cure. Predicting the disease early helps cure the patient with simplicity. Data mining techniques is a fast-growing technique used for gathering significant data and predicting the output. Necessary data from the patient like age, Chest Pain Type, Blood Sugar level etc, is composed and classified using data mining technique and the disease is simply identified. Thus necessary treatment can be taken at an early stage decreasing the rate of death. Main risk factors include diabetes, abdominal obesity, hypertension, etc. Coronary is a heart disease that is caused due to incorrect beating of the heart, whether asymmetrical, too fast or too slow. It happens when voltaic impulses in the heart don't work correctly. This disease does not have any obvious symptoms and hence it hard to predict [2].

The medical data mining is a very important research field due to its significance in the development of several applications in flourishing healthcare domain [3]. While summarizing the deaths occurring worldwide, the heart disease appears to be the important cause. The identification of the risk of heart disease in a person is complex task for medical practitioners because it requires years of experience and intense medical tests to be conducted. In this research work, we used many parameters to predict heart disease.

II. Background And Literature Review

Rising number of heart patients worldwide have motivated researchers to do comprehensive research to expose unknown patterns in clinical datasets. The Cleveland and Framingham dataset are used here, which are available on UCI and Kaggle repository. The classification goal of this research is to predict whether the patient, a) having heart disease or not (in Cleveland dataset) and b) has a 10-year risk of future heart diseases (in Framingham dataset). The data analysis is carried out in Python which is powerful data science application software.

S. Sharanyaa et al., 2020, In this research work, machine learning will expect the heart disease as well as non-functional state of the heart by the required clinical data value. In classification method, the entire dataset is separated into 70% of data for training and 30%of data for testing. The prediction of heart disease is based on machine learning algorithms like k-nearest neighbor algorithm (KNN), support vector algorithm (SVM), Decision tress algorithm, Random forest (RF) algorithm. The proposed hybrid approach is used to syndicate the characteristics of fuzzy logic and k-nearest neighbor algorithm which offers 94% accuracy [4].

JaishriPandhariWankhede et al., 2021, The purpose of the work is to throw light on few prevailing heart disease predicting methods as well as proposed a Hybrid Random Forest Model Joined with Linear Model (HRFMILM) for predicting also detecting the HDs at an initial stage. Averaging in Random Forest Model (RFM) increases the complete accuracy and decreases the probability of overfitting. Experimental outcomes established that the combination of Linear Model with RFM makes the modest estimation procedure with developed complete accuracy than the respective models. Further, the proposed method relates the prediction performance of few prevailing approaches in terms of parameters, precision, namely, recall and F1-score [5].

Pabitra Kumar Bhunia et al., 2021, In this research work, dissimilar classifiers were examined by performance assessment to classify the Heart Disease dataset to classify it appropriately and or to Predict Heart Disease cases with least attributes. In this case, a Heart Disease Prediction System is developed by using the algorithms Logistic Regression, Decision Tree, K Nearest Neighbor, Random Forest Classifier, and Support Vector Machine algorithms to predict the heart disease hazard level. The outcomes expose that the Random Forest Classifier and Support Vector Machine attained the maximum accuracy of 90.32%, whereas 87.09%, 70.96%, and 83.87% accuracy scores are attained by logistic regression, KNN classifier, and decision tree correspondingly [6].

U. Sivaji et al., 2021, The project focuses mostly on cardiovascular disease prediction in the real world. Heart disease prediction includes numerous risk factors. They proposed a new strategy in this article that targets to detect significant characteristics by applying AI techniques to improve the cardiovascular expectation accuracy. The preliminary model offers a range of highlights and number of well-known methods of classification. By the coronary disease expectation model with a hybrid random forest linear model that is amalgam of two dissimilar algorithms. They produced an improved presentation level with an accuracy level of 88.7 percent. Dataset is from the UCI repository. The proposed HRFLM algorithm will help doctors in diagnosis heart patients [7].

Vanitha Guda et al., 2020, In this work, the authors had proposed a method that targets at finding important features by applying machine learning techniques resulting in improving the accuracy in the prediction of cardiovascular

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Issue 12, December 2021

disease. In this work, they produced an improved performance level with great accuracy over the prediction model for heart disease by the hybrid technique. The key objective of the work is to increase the performance accuracy of heart disease prediction by uniting two models which gives better accuracy. This work presents comparative outcomes of numerous models based on machine learning and deep learning algorithms and techniques and analyses their performance [8].

III. Objectives Of The Research

Data mining techniques on medical dataset support in the prediction of heart disease with the help of Heart disease prediction system. The system has the capability to discover and extract unknown knowledge associate with heart disease from historical obtainable dataset.

The main purposes of this research work are given below:

1) To improve an intelligent clinical decision support system for the prediction of heart disease.
2) To assemble and generate dataset on the possibilities of heart disease (Cleveland and Framingham Datasets for heart risk assessment are used).
3) To simplify the data composed for computer analysis by pre-processing.
4) Applying hybrid approach (KNN & Random Forest) to predict heart disease in patient.
5) Compare the result of proposed method with the classification result of Naïve Bayes, SVM, KNN, Decision Tree, Random forest, Logistic Regression based on accuracy, precision, recall, f-measure, sensitivity and specificity.
6) To validate the proposed method by performance analysis based on accurateness level.

IV. Proposed Methodology

The proposed hybrid approach has 5 steps which shown in Figure.1. The steps in proposed methodology are

1. Data collection
2. Data Pre-Processing
3. Training and Testing the Datasets
4. Implementing Hybrid Approach (KNN & Random Forest)
5. Classification Result based on Accuracy
6. Compare result with Decision Tree, KNN, Logistic Regression, Naive Bayes and Random Forest.

1) Data Collection

In this research, two datasets are used: a) Cleveland dataset which is obtained from UCI repository and b) Framingham dataset which is obtained from Kaggle repository. The two datasets used in this research are discussed below:

(i) Cleveland Dataset

The Heart Disease database dataset contains 76 attributes from UCI Machine Learning Repository. Subset of 14 of 76 attributes is utilized for experimental purposes. The major goal is to predict the presence of heart disease in patients which is valued from 0 to 4. Cleveland database experimented results distinguish the presence of heart disease with values 1,2,3,4 and absence of heart disease with the value 0. The 14 attributes used are described below.

Cleveland Dataset Attributes

1. age: Age of the patients(mentioned in years)
2. sex: Gender of the patient
   - *Value 1: male
   - *Value 0: female
3. cp: Type of Chest pain,
   - *Value 1: typical angina
   - *Value 2: atypical angina
   - *Value 3: non-anginal pain
   - *Value 4: asymptomatic
4. trestbps: Resting of blood pressure (in mm Hg)
5. chol: Serum cholestoral in mg/dl
6. fbs: Fasting blood sugar > 120 mg/dl?
   - *Value 1: true
   - *Value 0: false
7. restecg: Resting electrocardiographic results
   - *Value 0: normal
   - *Value 1: having abnormality in ST-T wave (inversions in T wave and/or elevation in ST or depression > 0.05 mV)
   - *Value 2: showing possible or specific left ventricular hypertrophy by Estes' criteria
8. thalach: Maximum heart rate achieved
9. exang: Chest pain(angina) after exercise?
   - *Value 1: yes
   - *Value 0: no
10. oldpeak: Depression of ST induced by the exercise relative to rest.
11. slope: Measure of slope for peak exercise.
    - *Value 1: up sloping.
    - *Value 2: flat.
    - *Value 3: down sloping.

12.ca: Number of major vessels colored by fluoroscopy. Attribute values ranges 0 to 3.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Issue 12, December 2021

13. thal: Not described
    *Value 3=normal
    *Value 6=treated defect
    *Value 7=reversible defect
14. num: Heart disease diagnosis (status of the angiographic disease)
    *Value 0: less than 50% narrowing of coronary arteries(no heart disease)
    *Value 1,2,3,4: > 50% narrowing. The value indicates the stage of heart diseases.

(ii) Framingham Dataset

The dataset used in this research is taken from Kaggle repository. Dataset consists of 500 instances. 15 attributes and 1 target variable. 'TenYearCHD' attribute is our target variable where the risk of heart disease is predicted as value (1, 0). The dataset is described below.

Framingham Dataset Attributes

a) Demographic risk factors:

1. `sex`: male or female
   *Value 1: male
   *Value 0: female
2. `age`: age of the patient
   *Value: Contains ages of our patients in whole numbers
3. `education`: education level of the patient
   *Value 1: for some high school
   *Value 2: for a high school diploma or GED
   *Value 3: for some college or vocational school
   *Value 4: for a college degree

b) Behavioral risk factors that associated with smoking:

4. `currentSmoker`: is the patient is a current smoker?
   *Value 1: Yes
   *Value 2: No
5. `cigsPerDay`: the number of cigarettes in a day.
   *Value 1: Contains Number of cigarettes smoked on average in one day in whole numbers

c) Patient Medical history risk factors:

6. `BPMeds`: is patient on blood pressure medication?
   *Value 1: Yes
   *Value 2: No
7. `prevalentStroke`: is the patient had stroke previously
   *Value 1: Yes
   *Value 2: No
8. `prevalentHyp`: is the patient was hypertensive
   *Value 1: Yes
   *Value 2: No
9. `diabetes`: whether or not the patient had diabetes
   *Value 1: Yes
   *Value 2: No

d) Risk factors from physical examination of the patient:

10. `totChol`: total cholesterol level
    *Value: Contains patient's total cholesterol level
11. `sysBP`: systolic blood pressure
    *Value: Contains patient's systolic BP level
12. `diaBP`: diastolic blood pressure
    *Value: Contains patient's diastolic BP level
13. `BMI`: Body Mass Index
    *Value: Contains patient's body mass index
14. `heartRate`: heart rate
    *Value: Contains patient's average heart rate
15. `glucose`: glucose level
    *Value: Contains patient's glucose level
16. `TenYearCHD`:10 year risk of coronary heart disease CHD (Target Variable)
    *Value 1: Yes
    *Value 2: No

2) Data Pre-Processing

Data mining technique that includes transformation of raw data into clear format is termed as Data pre-processing. Data pre-processing makes raw data for further processing. The data pre-processing method is reactive, as it is assumed that the data is ready for analysis. The key difficulty for data pre-processing is the inconsistency among data sets.

a) Data Cleaning: The process of filling in missing values, smoothen noisy data, detect & remove outliners and resolution of inconsistencies is termed as Data cleaning.

b) Data Integration: Integration of multiple databases, data cubes, or files is termed as Data integration.

c) Data Transformation: Data normalization & aggregation is termed as Data transformation

d) Data Reduction: Process of abridged representation in volume but produces same or similar analytical results.

*e) Data Discretization:*Portion of data reduction but with particular significance, specifically for numerical data.

3) Proposed Hybrid of KNN And Random Forest (HKNNRF) Technique

The proposed hybrid approach is used combining the characteristics of K-Nearest Neighbor (KNN) and Random Forest. There are three steps in performing the hybrid technique:

1. Determine the output probabilities of each model
2. Finding the optimized weight

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Issue 12, December 2021

3. Applying combined model (KNN and Random Forest) for heart disease prediction with the help of weighted average

The proposed model initially finds the probabilities of the output for each model by using the pred_proba which give the probabilities in array format. Then, using log loss function, the optimized weight is determined, which combines the two classification models which has low classification error rate. At last, heart disease will be predicted using two classification models KNN and Random Forest using weighted average. As the experimental outcome, the result of proposed hybrid approach is compared with the results of machine learning classification algorithms: Decision Tree, KNN, Random Forest, Naïve Bayes and logistic Regression.

4) Classification Algorithms

The following classifiers are applied on the resulting dataset of the k-means clustering. Based accuracy level of classification, top classifier will be acquired among the following [9].

1. *Naive Bayes:*The idea of independence among every pair of features is on the basis of Bayes' theorem which is used in this algorithm. Naive Bayes classifiers works well in day to day circumstances like classification of documents and spam filtering.
2. *K-Nearest Neighbors:* This classification is usually termed as lazy learning as it does not try to create a shared internal model but the examples of training data are stored. Majority vote of the k nearest neighbors are calculated at each point in this classification.
3. *Decision Tree:* Given data attributes contained with classes, are categorized with a set of rules produced by decision tree.
4. *Random Forest:* This classifier positions number of decision tree on sub samples of given dataset. Accuracy of the model and over-fitment is predicted by average rise on the decision trees. Sub samples size is same as the original example size but with replacement.
5. *Logistic Regression:* Logistic Regression is a supervised Machine Learning algorithms used for classification i.e. to predict discrete valued outcome. It is a statistical approach that is used to predict the result of a dependent variable based on observations given in the training set.

V. EXPERIMENTAL RESULT

During the Modeling Phase, the algorithms for building models that would classify the students into the two classes – with and without heart disease, depending on their heart disease related medical data, are considered and selected. Here, a hybrid approach is applied using K-Nearest Neighbor and Random forest classification algorithms. Initially the dataset is pre-processing to avoid missing data. Then the dataset is analyzed using Hybrid of KNN and Random Forest approach and its result is compared with the result of machine learning classifiers Naïve Bayes, K-Nearest Neighbor, Decision Tree, Random Forest, and Logistic Regression. These classification algorithms are selected because they are frequently used algorithms for research purposes and have potential to achieve good outcome. Moreover different approaches are used for generating the classification models, which increases the probability for finding a prediction model with high classification accuracy. The experimental research is done by using Python. The sklearn library in python contains efficient tools for machine learning and statistical modeling including classification, and clustering. In The result generated by the classification algorithms in Scikit-learn, are compared by using the evaluation measures shown in following Table 1.

Table 1. Classification Evaluation Measures

Matrix	Description	Calculation
Precision	The classifier ability not to label as positive a sample that is negative.	Precision = (TP)/ (TP+FP)
Recall	The classifier ability to find all the positive samples.	Recall = (TP)/ (TP+FN)
F1 score	It is the harmonic mean of precision and recall which is better than accuracy.	F1_score = (2 * Precision * Recall) / (Precision + Recall)
Support	The number of occurrences of each class in the 'y_true'	
MacroAvg	Macro average is the mean average precision /recall/F1 of all classes.	macroavg = (precision of class0 + precision of class1)/2
Weightedavg	The total number true positive of all classes/total number of	Weightedavg = TP(of all classes)/total

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Issue 12, December 2021

	objects in all classes.	number of objects in all classes

The outcome of this research is found out in term of accuracy of the classification algorithm. The accuracy has been found with the help of confusion matrix. The Accuracy, Sensitivity and Specificity of the algorithm is calculated using the following formula:

Accuracy = {(TP + TN) / TP + FP + TN + FN)} * 100

Sensitivity = TP / (TP+FN)

Specificity = TN / (TN+ FP)

The experimental result of Hybrid approach in Cleveland and Framingham datasets are discussed below:

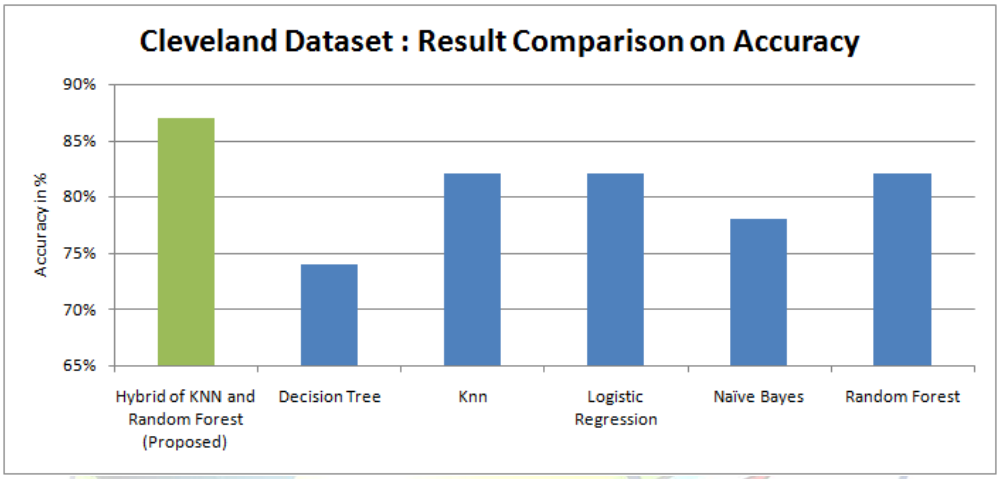

Figure 1. Cleveland Dataset: Result comparison

Table 2. Cleveland Dataset: Result of various models with proposed model

Algorithms	Accuracy (in %)	Accuracy	Precision	Recall	F-measure	Sensitivity	Specificity
Hybrid of KNN and Random Forest	87%	0.87	0.50	0.05	0.10	0.99	0.05
Decision Tree	74%	0.74	0.69	0.71	0.70	0.76	0.71
KNN	82%	0.82	0.77	0.83	0.80	0.81	0.83
Logistic Regression	82%	0.82	0.33	0.05	0.08	0.98	0.05
Naïve Bayes	78%	0.78	0.31	0.24	0.27	0.89	0.24
Random Forest	82%	0.82	0.86	0.71	0.78	0.90	0.71

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Issue 12, December 2021

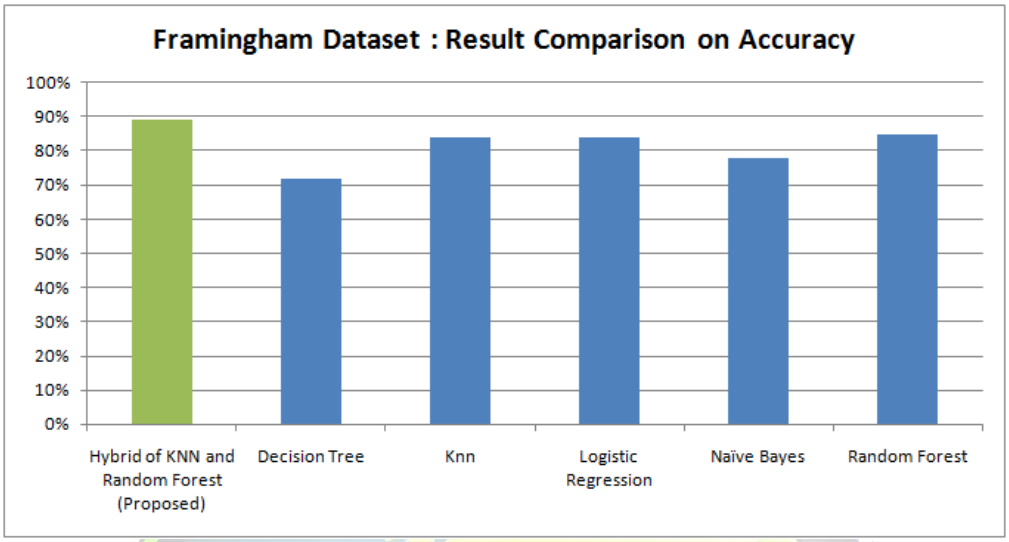

Figure 2. Framingham Dataset: Result comparison

Table 3. Framingham Dataset: Result of various models with proposed model

Algorithms	Accuracy (in %)	Accuracy	Precision	Recall	F-measure	Sensitivity	Specificity
Hybrid of KNN and Random Forest	89%	0.89	0.67	0.08	0.15	0.99	0.08
Decision Tree	72%	0.72	0.18	0.38	0.24	0.77	0.38
Knn	84%	0.84	0.23	0.12	0.16	0.94	0.12
Logistic Regression	84%	0.84	0.18	0.08	0.11	0.95	0.08
Naïve Bayes	78%	0.78	0.31	0.24	0.27	0.89	0.24
Random Forest	85%	0.85	0.25	0.12	0.17	0.95	0.12

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Issue 12, December 2021

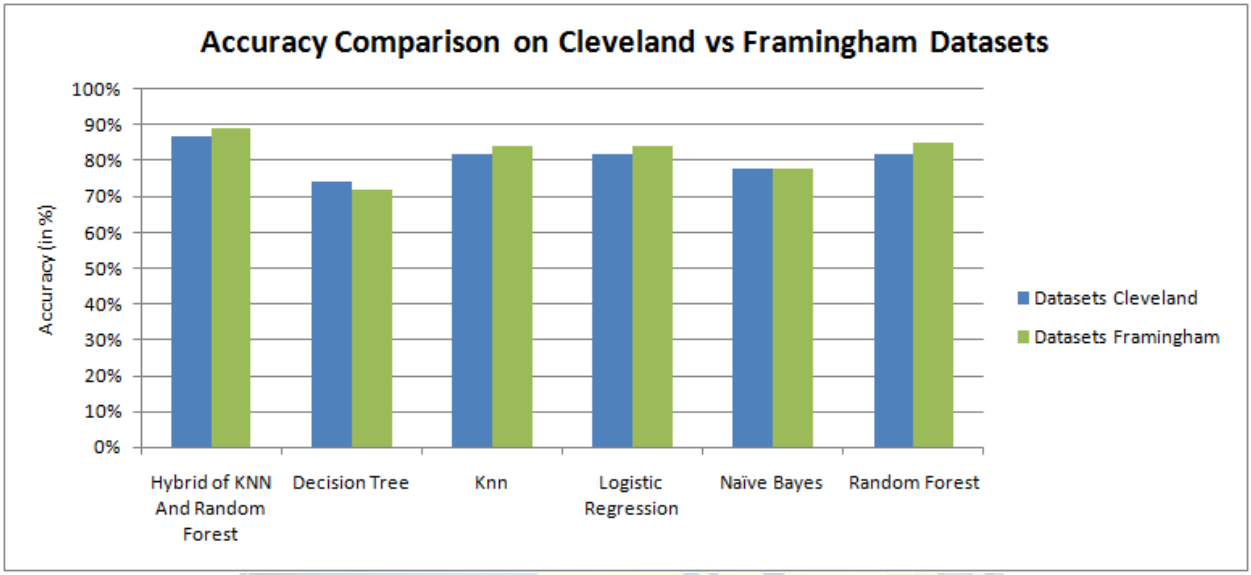

Figure 3. Accuracy Comparison on Cleveland and Framingham Datasets

Table 4. Accuracy Level (Result)

Classifier	***Datasets***	
	Cleveland	***Framingham***
Hybrid of KNN And Random Forest	87%	89%
Decision Tree	74%	72%
Knn	82%	84%
Logistic Regression	82%	84%
Naïve Bayes	78%	78%
Random Forest	82%	85%

From the above Table 2 and Figure 1, it is clearly shown that the accuracy rate obtained from the proposed experimental research using hybrid approach on Cleveland dataset resulted in highest accuracy and from the above Table 3 and Figure 2 shows the accuracy rate obtained from the proposed experimental research using hybrid approach on Framingham dataset resulted in highest accuracy. The Accuracy obtained from the Naïve Bayes, KNN, Decision Tree, Random forest, Logistic Regression and Hybrid of KNN and Random Forest classifiers on Cleveland and Framingham Datasets are stated in the above table 4. From the experimental result, the proposed approach on both the datasets results is better when compared to other machine learning algorithms in terms of accuracy.

VI. CONCLUSION

Heart disease is one of the more commonly prevailing diseases as per statistics in the world. Heart disease is an intricate disease and lots of lives are lost every year. Early prediction could help with appropriate treatment to keep it under control. If unnoticed, patient might end up with short span of life. In this research work, we have proposed a hybrid of K-Nearest Neighbor (KNN) and Random Forest algorithms. Cleveland & Framingham Heart disease datasets are collected from UCI and Kaggle Repository are taken for

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Issue 12, December 2021

analysis. In our Hybrid approach, the KNN and Random Forest algorithms are combined and applied to the two different dataset after pre-processing & training the dataset. The outcome of the hybrid approach is compared with the result of traditional 5 machine learning algorithms: Naïve Bayes, Logistic Regression, KNN, Decision Tree and Random Forest for analyzing the better the result in two different datsets. Out of the 5 machine learning classification algorithms, our proposed hybrid of KNN and Decision Tree algorithm has highest predictability in predicting heart disease. As the outcome of this research work, Hybrid of KNN and Decision Tree algorithms provides more accurate result as 87% for Cleveland dataset and 89% for Framingham dataset which is the highest accuracy compared to other algorithms. This will help medical physicians to make decision on heart disease occurrence with patients at an early stage with accuracy. As a future enhancement, data from various countries can be collected and create an accurate predictive model for heart disease diagnosis. Also we can predict the heart disease by considering more attributes that are related to the human life style that causes heart disease.

REFERENCES

1. K. Helini and P. Rajendra Prasad, "Hybrid System to prediction of Heart Disease using Data Mining Naive Bayes algorithm", International Journal of Scientific Research in Computer Science Applications and Management Studies, Vol. 9, Issue. 2, March 2020.
2. Pavithra M, Sindhana. A. M, Subajanaki T and Mahalakshmi S, "Effective Heart Disease Prediction Systems Using Data Mining Techniques", Vol. 25, Issue. 3, Annals of the Romanian Society for Cell Biology, March 2021.
3. H. Benjamin Fredrick David and S. Antony Belcy, "Heart Disease Prediction Using Data Mining Techniques", ICTACT Journal on Soft Computing, Vol. 9, Issue. 1, October 2018.
4. S. Sharanyaa, S. Lavanya, M. R. Chandhini, R. Bharathi and K. Madhulekha, "Hybrid Machine Learning Techniques for Heart Disease Prediction", International Journal of Advanced Engineering Research and Science (IJAERS), Vol. 7, Issue. 3, 2020.
5. Jaishri Pandhari Wankhede, Palaniappan S and Magesh Kumar S, "Heart Disease Prediction Using Hybrid Random Forest Model Integrated with Linear Model", Advances in Parallel Computing Technologies and Applications, pp. 370-376, 2021.
6. Pabitra Kumar Bhunia, Arijit Debnath and Poulami Mondal , "Heart Disease Prediction using Machine Learning", International Journal of Engineering Research & Technology (IJERT), Vol. 9, Issue. 11, 2021.
7. U. Sivaji, N V Krishna Rao and Chinnareddy Srivani, "A Hybrid Random Forest Linear Model approach to predict the Heart Disease", Annals of R.S.C.B., Vol. 25, Issue. 6, 2021.
8. Vanitha Guda, Shalini K, Shivani C, "Heart Disease Prediction Using Hybrid Technique", Journal of Interdisciplinary Cycle Research, Vol. XII, Issue. VI, 2020.
9. G. Jignesh Chowdary, Suganya. G and Premalatha. M, "Effective Prediction of Cardiovascular Disease using Cluster of Machine Learning Algorithms", Journal of Critical Reviews, Vol. 7, Issue. 18, 2020.
10. R. Indrakumari , T. Poongodi , Soumya Ranjan Jena, "Heart Disease Prediction using Exploratory Data Analysis", International Conference on Smart Sustainable Intelligent Computing and Applications under ICITETM2020, pp. 130–139, 2020.
11. Ruban, Vivek and Krithi, "Heart Disease Prediction using Machine Learning Models", International Journal of Recent Technology and Engineering (IJRTE), Vol. 8, Issue. 5S, Jan 2020.
12. Apurv Garg, Bhartendu Sharma and Rijwan Khan, "Heart disease prediction using machine learning techniques", IOP Conf. Series: Materials Science and Engineering, 2021.
13. Safa Sadat, Anish Xavier, and Sachin Agustine, "Cardiac Disease Recognition through Data Mining Techniques", International Research Journal of Modernization in Engineering Technology and Science, Vol. 02, Issue. 12, Dec 2020.
14. Galla Siva Sai Bindhika, Munaga Meghana, Manchuri Sathvika Reddy and Rajalakshmi, "Heart Disease Prediction Using Machine Learning Techniques", International Research Journal of Engineering and Technology (IRJET), Vol. 07 Issue. 04, Apr 2020.