International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 9 Issue: 1
DOI: https://doi.org/10.17762/ijfrcsce.v9i1.2126
***Article Received**: 26 November 2022 **Revised**: 14 January 2023 **Accepted**: 08 February 2023*

The Study and Efficacy of Conventional Machine Learning Strategies for Predicting Cardiovascular Disease

[1]Hamsitha Challagundla, [2]R S Sushanth, [3]Shivaji Potnuru
[1]Information Technology, Vellore Institute of Technology, Vellore, India.
Email: hamsitha.c@gmail.com
[2]Electronics and Communication Engineering, Vellore Institute of Technology, Vellore, India.
Email: sushanthraghava@gmail.com
[3]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India.
Email: shivajipotnuru12@gmail.com

Abstract: Regarding medical science, cardiovascular disease is the main cause of death. Testing patient samples for cardiac disease can save lives and lower mortality rates. During a subsequent visit, the right remedies should be outlined and prescribed. One of the most important factors in preemptive cardiac disease diagnosis is accuracy. Based on this factor, many research approaches were examined and compared. According to the analysis of these approaches, new procedures appear to be more advanced and reliable in detecting cardiac illness. A notation of the methods and their underlying themes and precision levels will be discussed. This paper surveys many models that use these methods and methodologies and evaluates their performance. Models created utilizing supervised learning methods, such as Support Vector Machines (SVM), K-Nearest Neighbour (KNN), Decision Trees (DT), Random Forest (RF), and Logistic Regression Units, are highly valued by researchers. For benchmark datasets like the Cleveland or Kaggle, the methodologies are derived from data mining, machine learning, deep learning, and other related techniques and technologies. The accuracy of the provided methods is graphically demonstrated.

Keywords: Cardiovascular disease, medical science, SVM, K-NN, random forests.

I. Introduction

The general structure of the body depends on the heart. The system is in charge of getting blood to particular areas of our bodies. If it doesn't work, the brain and other organs could shut down, which would cause death in a matter of minutes. Lifestyle changes, stress at work, and bad eating habits contribute to the increase of many heart-related illnesses. Heart conditions have overtaken all other causes of death as the notable top killer worldwide. The World Health Organization (WHO) reports that heart conditions are the thirty-first biggest cause of death globally, taking the lives of 17.7 million people each year [1]. In India, heart-related diseases also overtook all other causes of death. 1.07 million people died in India from cardiac diseases in 2016, according to the 2016 Global Burden of Disease Study, which was released on September 15, 2017. Heart-related illnesses raise a person's expenditures for medical care while also decreasing their productivity. According to estimates made by the WHO, Asian countries may have lost up to $237 billion between 2005 and 2015 due to heart-related or blood vessel disorders. Therefore, it is crucial to forecast heart-related disorders accurately[2].

Medical organizations from throughout the world gather information on a variety of issues relating to health. Using a variety of machine learning (ML) approaches, this knowledge is exploited to produce insightful results [3]. But, the amount of data gathered is staggering, and this knowledge will frequently be inconsistent. ML approaches are used to analyze these datasets, which are too big for human brains to process [4]. This has led to recent success in effectively predicting the presence or absence of heart-related illnesses using these algorithms. Research on ML has been used in multiple trials to treat various diseases, including skin and lung cancer [5, 6, 7].

Coronary heart disease is the most common worldwide (CHD). Some term it coronary artery disease (CAD). Lewy bodies in the circulatory system's capillaries and veins are a defining feature of this illness. As a direct result, the heart's internal organs do not receive the necessary blood and oxygen and have poor circulation. Most heart diseases are specified in Fig. 1; some are described below.

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 9 Issue: 1
DOI: https://doi.org/10.17762/ijfrcsce.v9i1.2126
***Article Received**: 26 November 2022 **Revised**: 14 January 2023 **Accepted**: 08 February 2023*

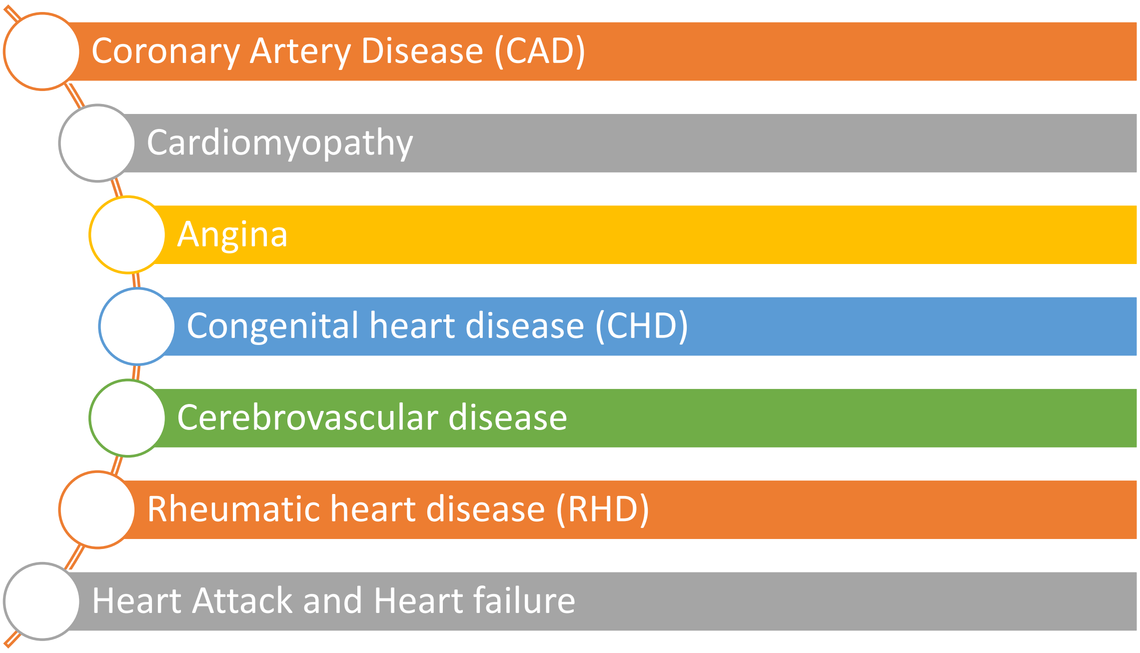

Fig. 1. Types of heart disease(s)

- **CAD** is the buildup of cholesterol plaque that causes the coronary artery to harden or narrow; (which supplies blood to the heart).
- **Cardiomyopathy** is a term used to describe various conditions that affect the heart muscle.
- Chest pain is caused by **angina**, a condition in which a heart muscle segment does not receive enough blood.
- **CHD** is a congenital disability of the heart's structure
- **Cerebrovascular disease** refers to a condition that affects the blood vessels that supply the brain**.**
- RHD, a rheumatic fever-related disorder that weakens the heart's muscles and valves.
- The section of cardiac muscle that loses blood supply after a **heart attack** dies permanently.
- A heart's pumping capacity decline causes **heart failure**.

This article will be organized as follows: Section II presents an overview of the main cardiac disease prediction approaches. Section III describes the results of our literature review on cardiac prognosis. In Section IV, we cover the traditional approaches to cardiology prognosis. In Section V, we see a comparison of several approaches to cardiovascular disease forecasting. The conclusion and potential future application are presented in Section VI.

II. Heart Disease Prediction Methods

Predicting cardiovascular diseases frequently uses data mining, machine learning, and deep learning.

Data Mining (DM)

DM is a technique for the organic development of information technology and is utilized in many technological, commercial, and scientific domains. The data sets for prediction techniques typically grow daily. With these massive data sets, retrieving the useful information concealed inside them is necessary. Since the importance of information is growing in today's competitive world [8],

Machine Learning (ML)

ML is a subfield of computer science concerned with automatic learning. Computing algorithms and statistics, which seek to conclude relationships between variables, foster its growth. There are unique computational challenges in developing statistical models from massive data sets, including billions or trillions of data points. Supervised and unsupervised learning processes can be implemented in computers [9].

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 9 Issue: 1
DOI: https://doi.org/10.17762/ijfrcsce.v9i1.2126
***Article Received**: 26 November 2022 **Revised**: 14 January 2023 **Accepted**: 08 February 2023*

Deep Learning (DL)

Through DL, the machine can invent novel ideas by combining existing ones. The DL method to ML was created with the help of our decade-long accumulation of artificial intelligence, statistical knowledge, and applied math skills. In recent years, DL's popularity and utility have skyrocketed because of its ability to train networks in depth. [10].

1.1 The motivation for the study

Various ailments afflict people all around the world. Nowadays, heart disease is a serious public health issue that claims both male and female victims. Heart disease claims 17.9 million lives per year, or 31% of all fatalities, according to the WHO. Heart problems can be predicted using machine learning tools and methodologies. Yet, no suitable model could forecast the disease more accurately or quickly enough. The effects of heart disease cannot yet be predicted or mitigated by any trustworthy automated technology. So, it would be a noteworthy achievement to use machine learning algorithms to lessen the disease's typical symptoms. Although greatly delaying the beginning of the disease, it has the potential to enhance the quality of life for cardiac patients. The deDevelopingrt disease prediction model for predicting the presence of heart disorders is the main motivation behind this research effort.

Also, as indicated earlier, this research intends to identify the classification system that will accurately forecast the illness. This study will be supported by comparing SVM, K-NN, DT, and RF algorithms for heart disease prediction. The most accurate algorithm is considered to be the best.

III. Literature Review

Much emphasis has been given to predicting cardiac disease utilizing machine learning, deep learning, and data mining techniques and approaches. The results thus far and future studies will be utilized to decide the most efficient methods for diagnosing cardiovascular illness. Researchers used a variety of datasets, algorithms, and procedures. Creating models to anticipate CVD diagnoses has grown significantly during the previous few decades. Automatic cardiovascular disease diagnosis and prediction is currently the most important medical topic. Effective treatment of heart disease depends on early identification. Researchers from all over the world have employed many techniques to identify heart disease early.

Several techniques have been applied to predict cardiac issues over the past 20 years. The most common DM techniques used in studies are SVM, Neural Networks, Regression, DT, etc. RF, Boosting, etc., used by ML. DL employs CNN, LSTM, and many other methods. Table 1 lists the relevant works on CVD in brief.

Table 1. Related works on CVD

Ref. and Year	**Dataset**	**Algorithm/ Method**	**Remarks**
Minas et.al., [11] (2010)	Public dataset	C4.5	Pruning is used to prevent overfitting. Using Laplace error estimation, the bottom-up pruning process is put into action. Substantial datasets are required for further research.
SreeHariRa o et.al.,[12] (2013)	STULONG	PSO	Due to a linear relationship between time and database size, the Predictive Risk Assessment of Atherosclerosis (PRAA) technique is numerically scaleable.
Roohallah et.al.,[13] (2013)	Personalized dataset.	SMO	SMO, NB, Bagging, and NN classification algorithms analyzed the dataset. These classifiers have 94.08% accuracy.
Jayshril et.al.,[14] (2014)	Cleveland	MLP	The stated accuracy is 98%. Compared to other procedures, this strategy yields good results.
Chaitanya et.al.,[15] 2015	Personalized dataset	Rule-based ML	The MEDICATION tag has the most annotations. Hence it caused the most errors in the test set with 90.7% accuracy.
Purushotta m et.al.,[16] (2016)	Cleveland	Decision Rules	The implementation uses KEEL. This approach tested at 86.3% accuracy.
Riccardo et.al.,[17] (2016)	Mount Sinai clinic	CNN	showed that DL can predict illnesses using EHRs.
Priyanga et.al.,[18] (2017)	Cleveland	NB	Naïve Bayes classifier can classify heart disease with more attributes. NBwa reports 86% accuracy.

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 9 Issue: 1
DOI: https://doi.org/10.17762/ijfrcsce.v9i1.2126
***Article Received**: 26 November 2022 **Revised**: 14 January 2023 **Accepted**: 08 February 2023*

MohdUsam a et.al.,[19] 2018	Dataset from china	RCNN	Pneumonia, coronary atherosclerosis, and heart disease are extracted here. 96.02% accuracy.
Beulah et.al.,[20] 2019	Cleveland	Ensemble	The best feature selection and large datasets improved performance
Anjan Nikhil Repak a et al, [21] (2019)	Personalized dataset	NB	The best feature selection and large datasets improved performance
Senthil Kumar Mohan et al. [22] (2019)	Personalized dataset	Decision Tree and K-NN	centered on the application of machine learning hybrid methods for accurate cardiac disease prediction with an accuracy of 88%
Binhua Wang et.al.,[23] (2019)	PLA hospital in china	DWNN	MT-DWNN has 0.9393 AUC, 0.01, better than other methods.
Shi et.al.,[24] (2020)	Renmin hospital in china	Statistics	The cardiac injury was an independent risk factor for in-hospital mortality among 416 consecutive COVID-19 patients.
Riyaz et.al., [25] (2021)	Personalized dataset	ANN C4.5	This study provides a comprehensive overview of several machine learning methods for accurate cardiac illness prognosis and treatment planning. The average prediction accuracy for ANN was 86.91 percent, while for C4.5's decision tree, it was just 74.0 percent.
Sekhar J et.al., [26] (2022)	UCI	TANFIS	The Internet of Things-based tuned adaptive neuro-fuzzy inference system (TANFIS) classifier predicts cardiac disease with 99.76% accuracy and 5.4% improvement.
EI Hasnony et.al., [27] (2022)	Cleveland	Active Learning	By combining user-expert feedback with sparsely labeled data, active learning approaches increase classification quality while lowering labeling costs.
El-Shafiey et al., [28] (2022)	Cleveland and Statlog dataset	PSO	To improve the precision of heart disease forecasting, we developed GAPSO-RF, a mix of genetic algorithms and PSO optimization. This technique achieves a 95% and 91% accuracy rate on two separate data sets.
Absar N et.al, [29] (2022)	Cleveland, Hungary, Switzerland, and Long Beach (CHSLB)	Numerous ML methods	Accuracy levels of 99.03%, 96.10%, 100%, and 100% were achieved across four machine learning models in this study for predicting cardiovascular disease.
Hassan D, et. al, [30] (2023)	Cleveland and another public dataset	PCA and ensemble methods	When applied to training and testing data, the suggested HD prediction method achieves accuracy ratings of 91.79 and 93.33 percent, respectively.
Ogundepo, E et al., [31] (2023)	Cleveland dataset	SVM	The support vector machine produced the best prediction performance with 85% accuracy, 82% sensitivity, 88% specificity, 87% precision, 91% area under the ROC curve, and 38% log loss value.
Sudha, V K et.al., [32] (2023)	Public dataset	LSTM	By integrating convolutional neural networks with a large short-term memory network, the suggested work improves the accuracy of conventional machine learning methods. Using the k-fold cross-validation method, the hybrid system was proven to be accurate 89% of the time.
Abdulsalam, G. et al., [33] (2023)	Cleveland	Bagging	With an accuracy of 90.16 percent, the Bagging-QSVC model is superior to the best available classifiers.

IV. Conventional Strategies

The experimental data were processed using the publicly available Anaconda 2020 version. Scientific computing typically employs the Python programming language (for machine learning and data science applications, preprocessing of massive amounts of data, predictive

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 9 Issue: 1
DOI: https://doi.org/10.17762/ijfrcsce.v9i1.2126
***Article Received**: 26 November 2022 **Revised**: 14 January 2023 **Accepted**: 08 February 2023*

analysis, and so on), and Anaconda is a free and conditional open-source distribution that seeks to simplify deployment and package management. In addition, Spyder is used as an integrated development environment, and Python (version 3.11.0) is used for programming and calculations. The dataset is split 70% for testing and 30% for training, and analysis is carried out with the help of the following methods.

K-NN

The test tuples are compared to similar training tuples in the relative neighbor classifier, a form of analogy learning. Training data is stored in an n-dimensional space, each tuple representing a point. The k-nearest neighbor classifier uses a set of k-training tuples to find the most similar pattern to an unknown tuple. In this case, the k sample tuples are the k nearest neighbors of the enigma tuple. Closeness can be measured with a distance metric like the Euclidean distance. The formula for determining the separation between two points is as follows:

$$\text{Distance } (X_1, X_2) = \sqrt{\Sigma}\,(x_1 i - x_2 i)^{2} \qquad \text{Eq. (1)}$$

SVM

This method of supervised machine learning can distinguish between linear and nonlinear data. SVM uses nonlinear mapping to change the original data. As a result, it looks for an ideal linear line to divide the hyperplane and a decision boundary to distinguish tuples of one class from those of another inside this new dimension.

Support vectors, crucial training tuples, and margins—also referred to as support vectors—are used by SVM to locate this hyperplane. The SVM's advantages include accuracy because of slow training times and the capacity to mimic sophisticated nonlinear decision limits. SVMs are much less likely than other algorithms to overfit. A hyperplane that divides the remaining examples of one class on one side from the remaining examples of the other class on the other is what SVM does best. A division line, commonly referred to as a hyperplane, can be expressed as follows:

$$Y = M.\,X + B \ = 0 \qquad \text{Eq. (2)}$$

M represents a weighted vector

B represents the bias

Decision Tree

Decision trees are supervised machine learning techniques where each branch indicates the result of a parameter test, and each leaf node contains a class label. The parent node, called the root node, is located at the top of the tree. A typical decision tree is displayed in Fig. 2. The best option can be chosen using a decision tree, which allows users to move from root to leaf to define a distinct class based on the most information gleaned. Decision trees can handle parameters that are continuous and constant. The main benefit of the decision tree is that it is prone to overfitting.

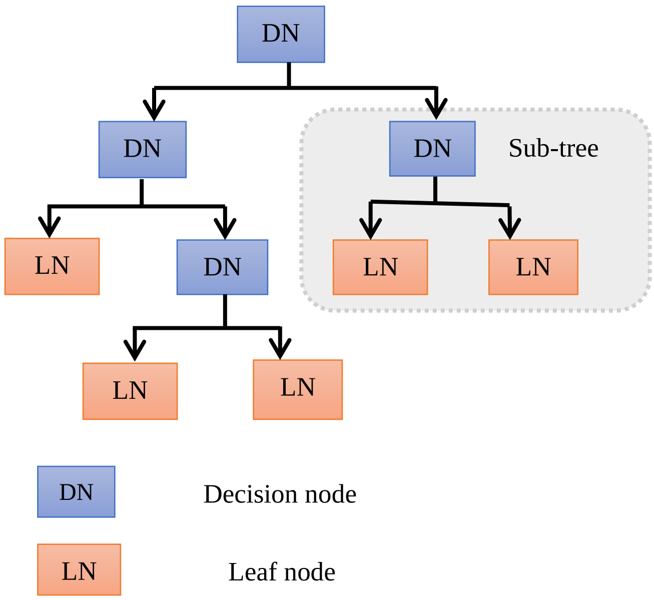

Fig. 2. Decision Tree

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 9 Issue: 1
DOI: https://doi.org/10.17762/ijfrcsce.v9i1.2126
***Article Received**: 26 November 2022 **Revised**: 14 January 2023 **Accepted**: 08 February 2023*

Random Forest

The online fitting method dynamically constructs the random forest's tree. Simply put, the bagging algorithm has changed. The individual trees are trained separately in a random forest, and the average projected values are added. The capacity of each tree in the forest and the connections between the trees affect the generalization error of a tree classifier. Although they can result in overfitting, random forests' main benefit is increased accuracy. As its name suggests, a random forest comprises several different decision trees that work together as an ensemble. Fig. 3 depicts our prediction method, which uses each tree in the random forest to forecast the class with the greatest number of options.

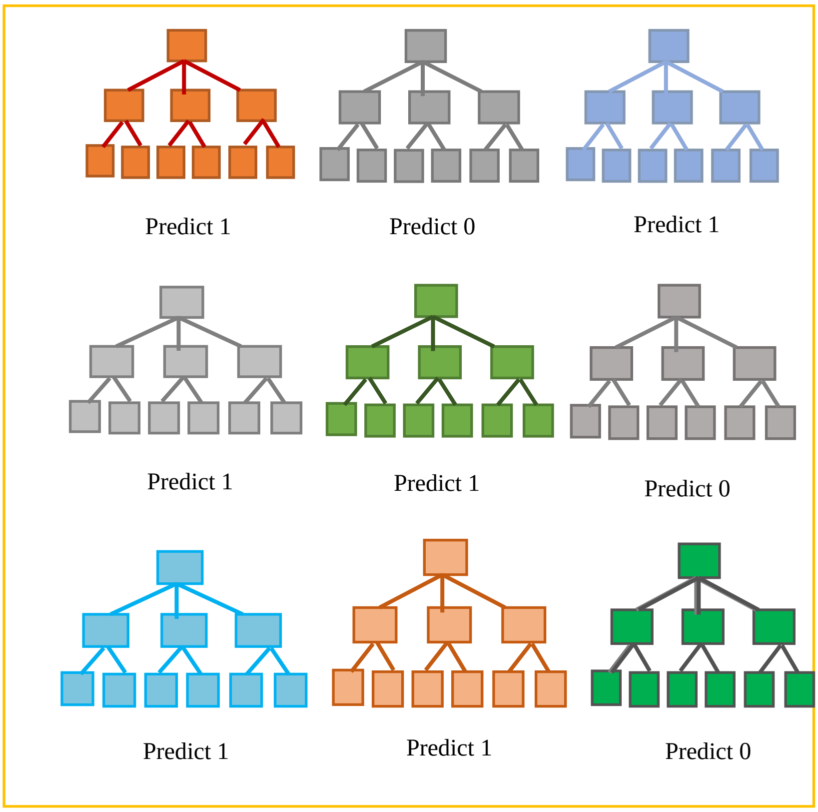

Fig. 3. Random Forests with multiple Decision Trees

V. Experimental Study

This section describes the dataset that was used, as well as the tests that we ran using various ML methods. The dataset was collected with an application from the UCI machine learning repository. Two datasets are available, one containing 303 instances and 14 attributes and the other containing 1026 instances and 14 attributes. The number of attributes in both datasets is the same. The first dataset contains the results of the experiments performed independently; the combined dataset's results are projected in this section. After combining the datasets, the final dataset will contain 1329 occurrences and 14 attributes. Table 2 provides a more detailed explanation of the features, and Table 3 displays the statistics associated with the dataset. Fig. 4 depicts the visualization of each attribute included in the dataset. A phrase used to describe how well these models work a confusion matrix. The following formulae calculate each model's final accuracy, precision, recall, and F1 score.

$$\text{Precision} = \frac{T_p}{T_p+F_p} \qquad \text{Eq. (3)}$$

$$\text{ecall}=\frac{T_p}{T_p+F_n} \qquad \text{Eq. (4)}$$

$$\text{F1 score} = 2 * \frac{Recall*Precision}{Recall+Precision} \qquad \text{Eq. (5)}$$

$$\text{Accuracy} = \frac{Number\ of\ exact\ predictions}{Total\ predictions} \qquad \text{Eq. (6)}$$

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 9 Issue: 1
DOI: https://doi.org/10.17762/ijfrcsce.v9i1.2126
***Article Received**: 26 November 2022 **Revised**: 14 January 2023 **Accepted**: 08 February 2023*

Table 2. Attributes of the UCI heart dataset

#	Column	Non-Null Count	Dtype
---	------	--------------	-----
0	age	303 non-null	int64
1	sex	303 non-null	int64
2	cp	303 non-null	int64
3	treetops	303 non-null	int64
4	chol	303 non-null	int64
5	fbs	303 non-null	int64
6	resting	303 non-null	int64
7	thalach	303 non-null	int64
8	exang	303 non-null	int64
9	oldpeak	303 non-null	float64
10	slope	303 non-null	int64
11	ca	303 non-null	int64
12	thal	303 non-null	int64
13	target	303 non-null	int64

Table 3. Statistics of the dataset

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
count	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000
mean	54.366337	0.683168	0.966997	131.623762	246.264026	0.148515	0.528053	149.646865	0.326733	1.039604	1.399340	0.729373	2.313531	0.544554
std	9.082101	0.466011	1.032052	17.538143	51.830751	0.356198	0.525860	22.905161	0.469794	1.161075	0.616226	1.022606	0.612277	0.498835
min	29.000000	0.000000	0.000000	94.000000	126.000000	0.000000	0.000000	71.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	47.500000	0.000000	0.000000	120.000000	211.000000	0.000000	0.000000	133.500000	0.000000	0.000000	1.000000	0.000000	2.000000	0.000000
50%	55.000000	1.000000	1.000000	130.000000	240.000000	0.000000	1.000000	153.000000	0.000000	0.800000	1.000000	0.000000	2.000000	1.000000
75%	61.000000	1.000000	2.000000	140.000000	274.500000	0.000000	1.000000	166.000000	1.000000	1.600000	2.000000	1.000000	3.000000	1.000000
max	77.000000	1.000000	3.000000	200.000000	564.000000	1.000000	2.000000	202.000000	1.000000	6.200000	2.000000	4.000000	3.000000	1.000000

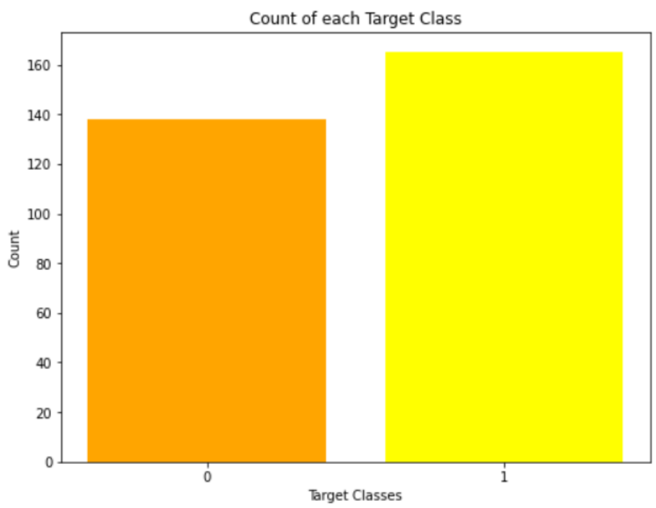

Fig. 4. The count of the target in the dataset

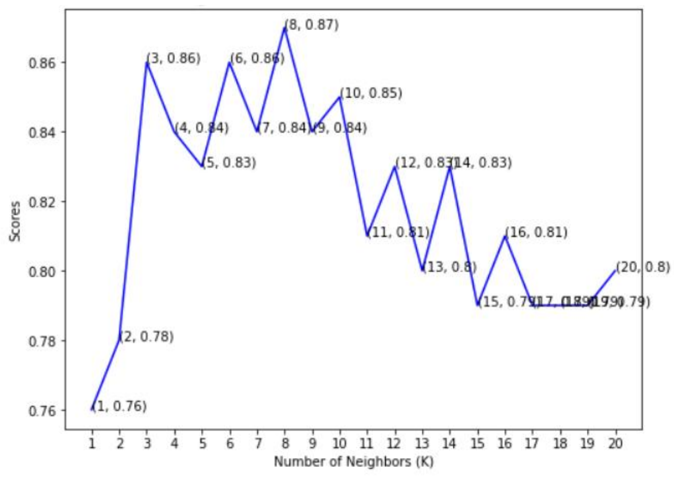

Fig. 5. K-NN scores for various K-values

The K-NN algorithm is applied to the dataset, and different k-values are tried in the experiment. The results are analyzed, and the plot is shown in Fig. 5. This classifier's accuracy is 87.0%.

Different kernels, linear, rbf, poly, and sigmoid, are tried for support vector machines, and the results are analyzed. The accuracy of these kernels is shown in Fig. 6, and the linear kernel gets the highest accuracy with 83.0%

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 9 Issue: 1
DOI: https://doi.org/10.17762/ijfrcsce.v9i1.2126
***Article Received**: 26 November 2022 **Revised**: 14 January 2023 **Accepted**: 08 February 2023*

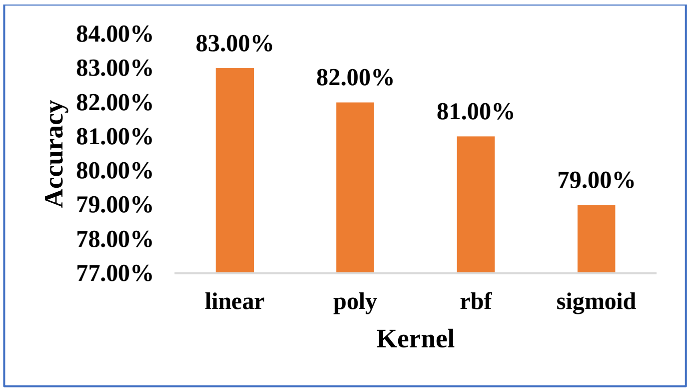

Fig. 6. The accuracy of SVM with different kernels

The decision tree model is trained on the same data set, and this model's result for a different number of top features is shown in Fig. 7.

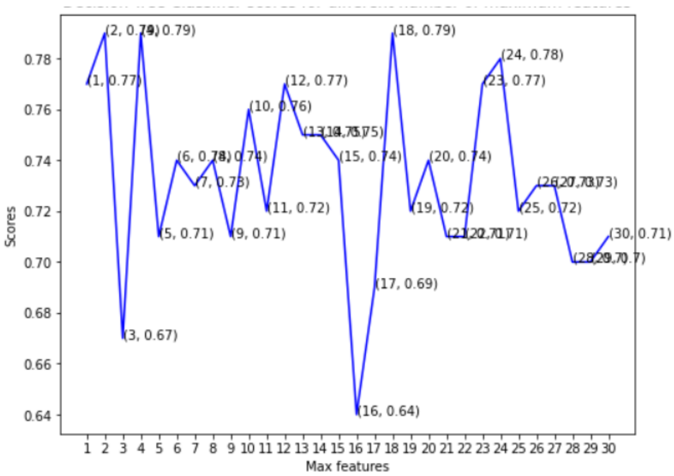

Fig. 7. Decision Tree Classifier scores for various numbers of maximum features

Finally, the random forest algorithm is trained, and the accuracy is noted for different estimators, i.e. (10, 100,200, 500, 1000). The results are shown in Fig. 8. The maximum accuracy is achieved with 100 estimators (84.0%).

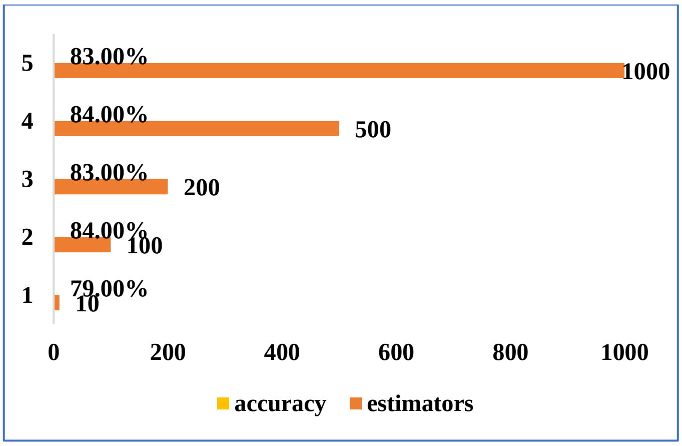

Fig. 8. Accuracy of random forest algorithm for different estimators

The experiments were performed on the combined data that had been preprocessed, and the results showed a significant improvement. Table 4 displays the better results obtained by each algorithm after applying preprocessing to the data.

Table 4. Evaluation measures of different classifiers on the UCI dataset

Classifier	Recall	Precision	Specificity	F1-score
K-NN	0.88	0.74	0.79	0.80
SVM	0.89	0.77	0.82	0.83
DT	0.96	1.0	1.0	0.98
RF	0.99	1.0	1.0	1.0

Table 5 below displays the efficacy of six machine learning algorithms to forecast heart illness. With an accuracy of 99.39%, the random forest outperforms the competition, followed by the decision tree at 97.59%, and the graphical representations are shown in Fig. 9.

Table 5. Accuracy of the ML algorithms on the UCI dataset

Algorithm	Accuracy
K-NN	82.05%
SVM	85.00%
DT	98.00%
Random Forest	99.00%

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 9 Issue: 1
DOI: https://doi.org/10.17762/ijfrcsce.v9i1.2126
***Article Received**: 26 November 2022 **Revised**: 14 January 2023 **Accepted**: 08 February 2023*

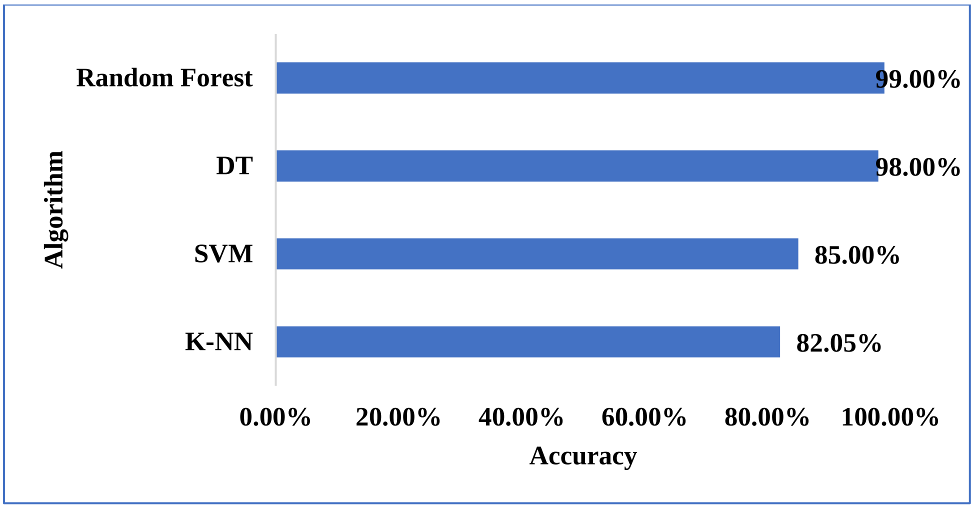

Fig. 9. Comparison results of ML algorithms on the UCI dataset

VI. Conclusion

The most useful contribution of this study was a comparison of different ML algorithms for early-stage CVD prediction. To begin, we have thoroughly examined the various strategies for predicting heart disease found in the body of published research. Following that, a case study will be conducted. Preprocessing techniques improved the dataset's quality by removing outliers and dealing with corrupted and missing values. In addition, we examine the outcomes of three distinct machine-learning algorithms that we utilize to forecast diseases. We know that it outperforms every other classifier on our dataset based on the results of our studies, with an accuracy of (100% for the training set) and (97.29% for the testing set). A 10-fold cross-validation technique was used to verify the robustness of K-NN, SVM, DT, and RF. Future work on this project could involve identifying exact knowledge patterns using deep learning, fuzzy logic, and other methods on vast source data.

REFERENCES

[1] Bani Hani, S. H., & Ahmad, M. M. (2023). Machine-learning Algorithms for Ischemic Heart Disease Prediction: A Systematic Review. *Current Cardiology Reviews*, *19*(1), 87-99.

[2] Chalapathi, M. M., Kumar, M. R., Sharma, N., & Shitharth, S. (2022). Ensemble Learning by High-Dimensional Acoustic Features for Emotion Recognition from Speech Audio Signals. *Security and Communication Networks*, *2022*.

[3] Cantinotti, M., Marchese, P., Giordano, R., Franchi, E., Assanta, N., Koestenberger, M., ... & McMahon, C. J. (2023). A review of echocardiographic scores for biventricular repair risk prediction of congenital heart disease with borderline left ventricle. *Heart Failure Reviews*, *28*(1), 63-76.

[4] Kalyani, B. J. D., Meena, K., Murali, E., Jayakumar, L., & Saravanan, D. (2023). Analysis of MRI brain tumor images using deep learning techniques. *Soft Computing*, 1-8.

[5] Ramana, K., Kumar, M. R., Sreenivasulu, K., Gadekallu, T. R., Bhatia, S., Agarwal, P., & Idrees, S. M. (2022). Early prediction of lung cancers using deep saliency capsule and pre-trained deep learning frameworks. *Frontiers in Oncology*, *12*.

[6] Meena, K., Veni, N. K., Deepapriya, B. S., Vardhini, P. H., Kalyani, B. J. D., & Sharmila, L. (2022). A Novel Method for Prediction of Skin Diseases Using Supervised Classification Techniques.

[7] Rudra Kumar, M., & Gunjan, V. K. (2022, May). Machine Learning Based Solutions for Human Resource Systems Management. In *ICCCE 2021: Proceedings of the 4th International Conference on Communications and Cyber Physical Engineering* (pp. 1239-1249). Singapore: Springer Nature Singapore.

[8] Kuruba, C., Pushpalatha, N., Ramu, G., Suneetha, I., Kumar, M. R., & Harish, P. (2022). Data mining and deep learning-based hybrid health care application. *Applied Nanoscience*, 1-7.

[9] Sun, H., & Pan, J. (2023). Heart Disease Prediction Using Machine Learning Algorithms with Self-Measurable Physical Condition Indicators. *Journal of Data Analysis and Information Processing*, *11*(1), 1-10.

[10] Hassan, D., Hussein, H . I., & Hassan, M. M. (2023). Heart disease prediction based on pre-trained deep neural networks combined with principal component analysis. *Biomedical Signal Processing and Control*, *79*, 104019.

[11] Amin, S. U., Agarwal, K., & Beg, R. (2013, April). Genetic neural network-based data mining in predicting heart disease using risk factors. In *2013 IEEE conference on information & communication technologies* (pp. 1227-1231). IEEE.

[12] Rao, V. S. H., & Kumar, M. N. (2012). Novel approaches for predicting risk factors of atherosclerosis. *IEEE journal of biomedical and health informatics*, *17*(1), 183-189.

International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE)
ISSN: 2454-4248 Volume: 9 Issue: 1
DOI: https://doi.org/10.17762/ijfrcsce.v9i1.2126
***Article Received**: 26 November 2022 **Revised**: 14 January 2023 **Accepted**: 08 February 2023*

[13] Alizadehsani, R., Habibi, J., Hosseini, M. J., Mashayekhi, H., Boghrati, R., Ghandeharioun, A., ... & Sani, Z. A. (2013). A data mining approach for diagnosis of coronary artery disease. *Computer methods and programs in biomedicine*, *111*(1), 52-61.
[14] Mirmozaffari, M., Alinezhad, A., & Gilanpour, A. (2017). Data mining classification algorithms for heart disease prediction. *Int'l Journal of Computing, Communications & Instrumentation Engg*, *4*(1), 11-15.
[15] Shivade, C., Malewadkar, P., Fosler-Lussier, E., & Lai, A. M. (2015). Comparison of UMLS terminologies to identify the risk of heart disease using clinical notes. *Journal of biomedical informatics*, *58*, S103-S110.
[16] Purushottam, Kanak Saxena, and Richa Sharma, "Efficient Heart Disease Prediction System," Elsevier, Procedia Computer Science, Vol. 85, pp. 962-969, 2016
[17] Riccardo Miotto, Li Li and Joel T. Dudley, "Deep Learning to Predict Patient Future Diseases from the Electronic Health Records," Springer European Conference on Information Retrieval, pp. 768–774, 2016
[18] P. Priyanga and N. C. Naveen, "Web Analytics Support System for Prediction of Heart Disease Using Naive Bayes Weighted Approach (NBwa)," 2017 Asia Modelling Symposium (AMS), Kota Kinabalu, pp. 21-26, 2017
[19] MohdUsama, Belal Ahmad, Jiafu Wan, M. ShamimHossain, Mohammed F Alhamid, and M. Anwar Hossain, “Deep Feature Learning for Disease Risk Assessment Based on Convolutional Neural Network With Intra-Layer Recurrent Connection by Using Hospital Big Data,” IEEE Access, Vol. 6, pp. 67927-67939, 2018
[20] C. Beulah Christalin Latha and S. Carolin Jeeva, "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques," Elsevier, Informatics in Medicine Unlocked, Vol. 16, 2019
[21] Anjan Nikhil Repaka, Sai Deepak Ravikanti, and Ramya G Franklin," Design And Implementing Heart Disease Prediction Using Naive Bayesian" IEEE Xplore Part Number: CFP19J32-ART; ISBN: 978-1-5386-9439-8, pp. 292-297, 2019
[22] Senthilkumar Mohan, Chandrasegar Thirumalai, and Gautam Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques" IEEE Access, DOI 10.1109/ACCESS.2019.2923707, pp. 81542-81554, 2019.
[23] Binhua Wang, Yongyi Bai, Zhenjie Yao, Jiangong Li, Wei Dong, Yanhui Tu, Wanguo Xue, Yaping Tian, Yifei Wang, and Kunlun He, "A Multi-Task Neural Network Architecture for Renal Dysfunction Prediction in Heart Failure Patients With Electronic Health Records," IEEE Access, Vol. 7, pp. 178392-178400, 2019
[24] Shi S, Qin M, and Shen B, "Association of cardiac injury with mortality in hospitalized patients with COVID-19 in Wuhan," JAMA Cardiol, Vol. 5, No. 7, pp. 802-810, 2020.
[25] Riyaz, L., Butt, M. A., Zaman, M., & Ayob, O. (2022). Heart disease prediction using machine learning techniques: a quantitative review. In *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021, Volume 3* (pp. 81-94). Springer Singapore.
[26] Sekar, J., Aruchamy, P., Sulaima Lebbe Abdul, H., Mohammed, A. S., & Khamuruddeen, S. (2022). An efficient clinical support system for heart disease prediction using TANFIS classifier. *Computational Intelligence*, *38*(2), 610-640.
[27] El-Hasnony, I. M., Elzeki, O. M., Alshehri, A., & Salem, H. (2022). Multi-label active learning-based machine learning model for heart disease prediction. *Sensors*, *22*(3), 1184.
[28] El-Shafiey, M. G., Hagag, A., El-Dahshan, E. S. A., & Ismail, M. A. (2022). A hybrid GA and PSO optimized approach for heart-disease prediction based on random forest. *Multimedia Tools and Applications*, *81*(13), 18155-18179.
[29] Absar, N., Das, E. K., Shoma, S. N., Khandaker, M. U., Miraz, M. H., Faruque, M. R. I., ... & Pathan, R. K. (2022, June). The efficacy of machine-learning-supported smart system for heart disease prediction. In *Healthcare* (Vol. 10, No. 6, p. 1137). MDPI.
[30] Hassan, D., Hussein, H. I., & Hassan, M. M. (2023). Heart disease prediction based on pre-trained deep neural networks combined with principal component analysis. *Biomedical Signal Processing and Control*, *79*, 104019.
[31] Ogundepo, E. A., & Yahya, W. B. (2023). Performance analysis of supervised classification models on heart disease prediction. *Innovations in Systems and Software Engineering*, 1-16.
[32] Sudha, V. K., & Kumar, D. (2023). Hybrid CNN and LSTM Network For Heart Disease Prediction. *SN Computer Science*, *4*(2), 172.
[33] Abdulsalam, G., Meshoul, S., & Shaiba, H. (2023). Explainable Heart Disease Prediction Using Ensemble-Quantum Machine Learning Approach. *Intelligent Automation & Soft Computing*, *36*(1).