Available online at www.qu.edu.iq/journalcm

JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS

ISSN:2521-3504(online) ISSN:2074-0204(print)

Machine Learning Techniques for Predication of Heart Diseases

Raed Hassan Laftah*[1]*, Karim Hashim Kraidi Al-Saedi*[2] **

Mustansiriyah University College of Science /Computer Department Iraq/ Baghdad

Email: raed.dirassat@gmail.com,dr.karim@uomustansiriyah.edu.iq

ARTICLE INFO

Article history:
Received: 14 /07/2024
Rrevised form: 25 /08/2024
Accepted : 29 /08/2024
Available online: 30 /09/2024

Keywords:

Heart disease

Data mining

Random forest

Support Vector Machine

Confusion matrices.

ABSTRACT

Heart disease, or cardiovascular illness, encompasses a wide range of disorders affecting the cardiovascular system. One of the trickiest things to do in medicine is to make predictions about cardiovascular disease. Nowadays, heart disease claims the life of almost one person every minute. Heart disease has several causes, but one of the most pressing issues is the lack of sensitive, precise methods for early identification, which makes proper management of the condition impossible. Automating the prediction process is necessary to prevent the hazards connected with cardiac disease diagnosis and to inform the patient at an early stage due to the intricacy of the condition. Data mining is extensively utilized in healthcare to forecast the occurrence of cardiovascular illness by analyzing massive and intricate medical records. To forecast cardiac problems, researchers conduct in-depth analyses of massive amounts of medical data using a wide range of data mining and machine learning algorithms. Here, we provide several heart disease characteristics and build a model using supervised learning methods like random forest and support vector machine (SVM). The Kaggle repository contains the cardiac condition dataset that is used in this research. Predicting patients' risk of heart disease is the main objective of this investigation. Confusion matrices were used for proposed system evaluation .The findings demonstrate that Random Forest achieves the highest level of accuracy, reaching 98.54 percent.

MSC..

https://doi.org/ 10.29304/jqcsm.2024.16.31646

1. Introduction

Heart disease and other cardiovascular diseases (CVDs), which account for a disproportionate share of all deaths worldwide, are the major causes of mortality. Cardiovascular disease is predicted to cause 17.9 million deaths per year [1] and 23.6 million deaths per year by 2030 [2] according to the World Health Organization. Heart disease, stroke, and other cardiovascular disorders account for a disproportionate share of fatalities in low- and middle-

∗Corresponding author

Email addresses:

Communicated by 'sub etitor'

income nations [3][4]. While several variables can increase the risk of HD, poor nutrition, smoking, sugar intake, and obesity are more common in high-income nations [5] [6]

Similar lifestyle variables and increased urbanization are also driving an increase in the incidence of chronic diseases in low- and middle-income nations [4].

. Interest in using these methods to foretell HD risk has been on the rise in recent years. Essential components of these forecasts are conditions such as diabetes, hypertension, irregular heart rates, and cholesterol levels [12]. Unfortunately, the accuracy of HD prediction might be affected when the provided medical data are inadequate.

Achieving exact predictions, particularly regarding the course of disorders like HD [14], is still challenging, but previous research have shown the applicability of data mining in disease prediction [13]. The main goal of this work is to improve medical diagnostics and preventative healthcare by applying these revolutionary data mining and ML algorithms to increase the accuracy of predicting the likelihood of HD.

This study aims to discover which patient variables are linked to a higher incidence of cardiovascular HDs, such as gender, age, kind of chest pain, and fasting blood sugar levels. In order to make this happen, we use a dataset that contains medical histories and characteristics of patients that is available publically on Kaggle and is stored in the UCI repository. Examining the relative accuracy of several ML models in predicting HD is the driving force behind this research.

The paper goes on to examine several ML algorithms individually [11] and then uses ensemble learning to improve prediction accuracy. Python code runs the whole thing on Google Colab, from data pretreatment to model calculation and outcome display. Using ensemble learning to fine-tune HD prediction accuracy, this all-encompassing method seeks to choose the best individual ML model.

2. Related Works

Medical cardiology is one area that has benefited greatly from the recent advances in data mining and ML. These methods have been successful in evaluating the ever-increasing amounts of medical data, as demonstrated in the cited publications, opening up new avenues for algorithm testing and development. Heart disease (HD) is an important topic for study since it is a major killer in underdeveloped nations. All of the linked articles stress how critical it is to find HD risk factors and symptoms early on. Early identification and prevention of HD, via the use of data mining and ML approaches, hold great promise for decreasing its impact on public health.

Shorewala, 2020: The authors in this line [4] devised a way to predict coronary heart disease using factor-based methods, advanced machine learning, and deep learning techniques. In this study, ensemble modeling techniques such as bagging and boosting are used to stack the basic classifiers. Their analyses using a 70,000-patient-record dataset showed accuracy of the models increasing with bagging by 1.96%, on average, and 2 to 4 percentage points

∗Corresponding author

Email addresses:

Communicated by ‘sub etitor’

in the overall model error rate with boosting. The ensemble models were observed to give an average of 73.4%, with the highest achieving AUC score of 0.73. The best performance is by stacking three classifiers; namely, KNN, RF classifier, and SVM, which results in 75.1%.These findings were also confirmed in an additional way by examining several data-analytic techniques, as well as K-Folds cross-validation to further attest to the model's performance.

Shimaa Ouf, 2021 [17], considered it as the objective to examine the efficiency of DM techniques coupled with four CKs in terms of HD prediction. These algorithms were implemented on the same eight classification data mining . The goal was to find the best combination which gives highest accuracy in predicting HD by accuracy, precision, recall and f-measure using datasets obtained from Kaggle and UCI machine learning repository. It was concluded that as a large dataset, with 70,000 records and using Neural Network met the highest potential accuracy which can reach up to 71.82% when it is applied together with holdout method and Repeated Random derived highest possible classification accuracy (89.01%) at small size data set of 303 records; Finally, the findings of this research would enable companies to detect HD among employees at an early stage that would contribute to increasing the overall productivity.

Dro˙zd˙z et al., 2022: ML algorithms were applied to detect new CVD risk factors among patients [18]. Blood biochemical evaluations and assessments of subclinical atherosclerosis were evaluated in 191 patients with MAFLD. The primary modeling consisted of multiple logistic regressions: a univariate filter method for ranking selected features, and principal component analysis (PCA). This model showed an accuracy of 85.11% for high-risk category and 79.17% for low-risk categories among CVD patients (1). The overall classification accuracy of the model (Area Under the Curve [AUC] value) was 0.87.The significant results suggest that machine learning could be leveraged to estimate the CVD risk only from clinically relevant factors in MAFLD patients.

Boukhatem, Youssef, & Nassif, 2022: They described their investigation to develop smart diagnostics using electronic health data to resolve the problem of cardiovascular disease diagnosis with high precision, as detailed in In[19]. The researchers established prediction models based on 4 classification algorithms: Multilayer Perceptron, Support Vector Machine, Random Forest and Naïve Bayes. This came after doing preprocessing and feature selection over the raw data, which was used as an input for this. . As per the performance metrics, SVM model had higher accuracy i.e., 91.67%.

Chandrasekhar & Peddakrishna,2023: This research [20] shows a new technique done to improve the prediction of heart disease that performed experiments with six different algorithms (RF, KNN, LR, NB classifier, Gradient Boosting and AdaBoost Classifier). In dealing with GridSearch CV and 5-Fold Cross Validation in the Cleveland and IEEE DataPort dataset to increase model accuracy. The Logistic Regression model with the Cleveland dataset had an accuracy of 90.16% and AdaBoost achieved an accuracy of 90% using the IEEE data set in this study. In addition, an advanced soft voting ensemble cannot reach its goal in accuracy; the Cleveland dataset flipped and yielded it to 93.44% due to blending all six algorithms [sic]. This improved approach also transferred to the IEEE DataPort dataset, reproducing similar same results with which were able to achieve an accuracy of 95% at max. Finally, the ensemble method exhibited improved results in comparison to any single algorithm used in isolation.

Khan et al, 2023: In their study [21], the researchers examined the use of prediction algorithms in early detection and precise prediction of CVD. Regarding patient diagnosis and therapeutic conditions in healthcare, whereby focus is a critical aspect .The data from these hospitals of Khyber Teaching Hospital and Lady Reading Hospitals in Pakistan was analyzed using DT, RF, LR, NB, and SVM methods have been used to treat the patients according to classified analysis for prediction of heart disease. Out of these algorithms, RF has proven to be the most effective for achieving 85.01% high accuracy in CVD prediction.

Bizimana et al., 2023: In this study [22], a novel prediction model, termed MLbPM ,was proposed, which aimed at predicting HD. Various techniques used in building this model include data scaling, optimal split ratios, parameter tuning and algorithm selection. To verify the performance of our model the experiments on UC Irvine HD datasets (used commonly to obtain

TABLE 1- RELATED WORKS

References	Model or Methods	Dataset	Results
Boukhatem, H. Y. Youssef, and A. B. Nassif, 2022	SVM and other classification algorithms after preprocessing and feature extraction	Electronic Health Records	Precision: 91.67%
Chandrasekhar and S. Peddakrishna,2023	Six algorithmic models, AdaBoost, LR, soft voting ensemble with GridSearchCV, and cross-validation	Cleveland dataset from IEEE Dataport	Cleveland: 93.44%, IEEE Dataport: 95%
Khan, M. Qureshi, M. Daniyal, K. Tawiah et al.,2023	DT, RF, LR, NB, SVM techniques	Patients data from Khyber and Lady Reading Hospitals	RF: 85.01%
Bizimana, Z. Zhang, M. Asim, A. El-Latif, A. Ahmed et al.,2023	LR algorithm within a ML-based prediction model (MLbPM) using data scaling and optimal settings	University of California Irvine HD dataset	96.7%

3. Material and Methods

Figure 1 shows the process flow for a Kaggle-based HD Prediction model. In order to comprehend the properties of the dataset and spot patterns or outliers, Exploratory Data Analysis (EDA) is the first step. In order to get the data ready for modeling, preprocessing activities like normalization are performed after EDA.

To help validate the models, the preprocessed dataset is divided into two parts: the training set and the testing set. A number of distinct models are trained: RF, and SVM. After learning from the training data, each model is evaluated for its ability to make predictions.

The component models are trained separately, used to provide the final output prediction. Because it may take use of the good parts of each model while reducing the bad, this method usually improves performance. We use a number of criteria to assess how well the ensemble model and the individual models perform.

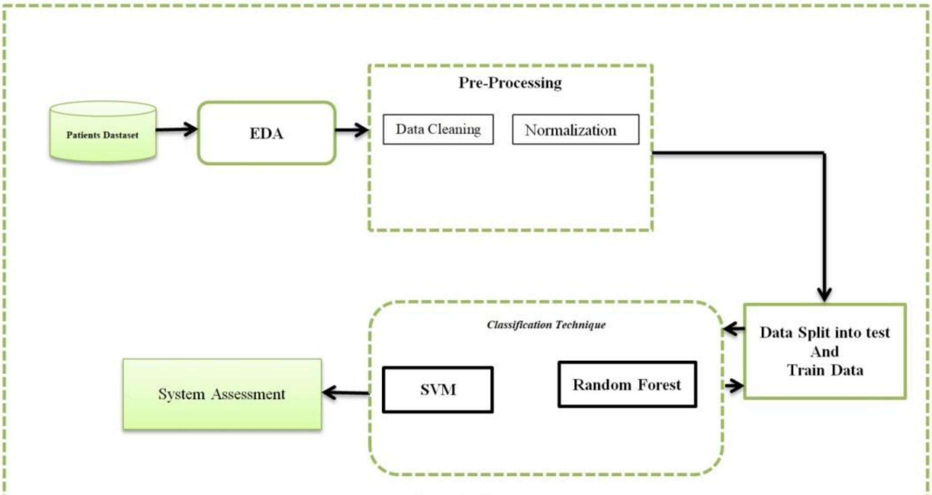
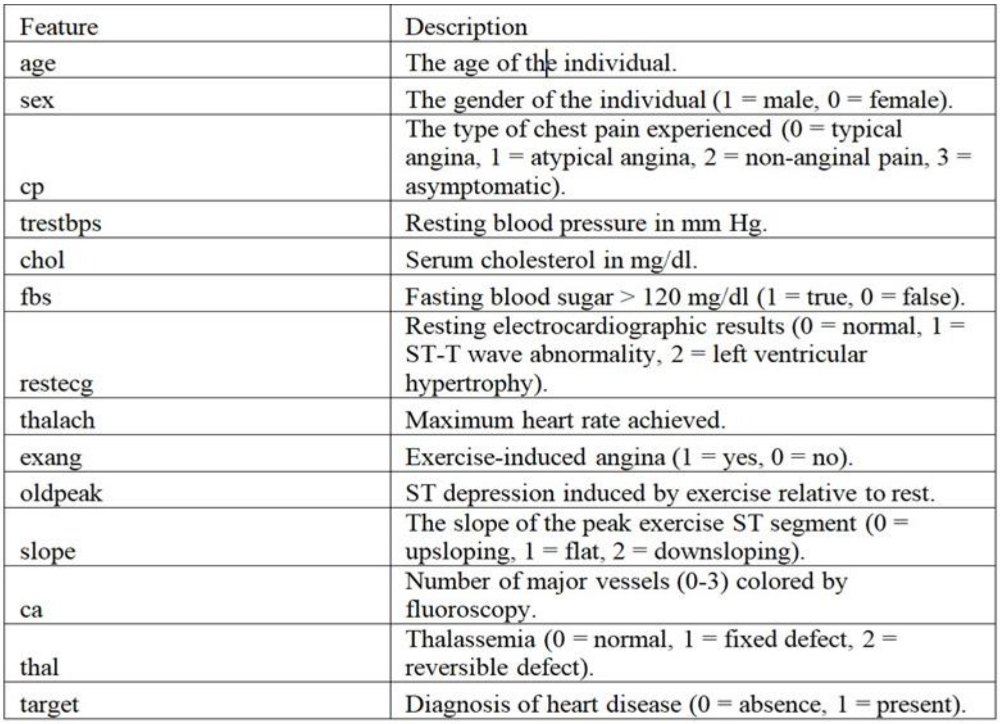


figure 1. HD system

3.1. Dataset Description

Our study's dataset is based on the famous Cleveland database housed in the UCI repository; an online version of this database is also accessible on Kaggle. Each of the 1025 examples has one of fourteen unique characteristics that are important for HD diagnosis. You may find a thorough explanation of each feature in Table 1. Factors such as age, sex, clinical and physiological indicators (such as different kinds of chest discomfort, resting blood pressure, serum cholesterol levels, and fasting blood sugar), and more are detailed in this table. Results from resting electrocardiographic tests, maximum heart rate, exercise-induced angina, exercise-induced ST depression relative to rest, slope of the peak exercise ST segment, number of major vessels colored by fluoroscopy, presence of thalassemia (a blood disorder affecting hemoglobin levels), and other variables are also part of this. The last property, "target," shows whether HD is present or not.

Table (1) :Dataset Features description



Feature	Description
age	The age of the individual.
sex	The gender of the individual (1 = male, 0 = female).
cp	The type of chest pain experienced (0 = typical angina, 1 = atypical angina, 2 = non-anginal pain, 3 = asymptomatic).
trestbps	Resting blood pressure in mm Hg.
chol	Serum cholesterol in mg/dl.
fbs	Fasting blood sugar > 120 mg/dl (1 = true, 0 = false).
restecg	Resting electrocardiographic results (0 = normal, 1 = ST-T wave abnormality, 2 = left ventricular hypertrophy).
thalach	Maximum heart rate achieved.
exang	Exercise-induced angina (1 = yes, 0 = no).
oldpeak	ST depression induced by exercise relative to rest.
slope	The slope of the peak exercise ST segment (0 = upsloping, 1 = flat, 2 = downsloping).
ca	Number of major vessels (0-3) colored by fluoroscopy.
thal	Thalassemia (0 = normal, 1 = fixed defect, 2 = reversible defect).
target	Diagnosis of heart disease (0 = absence, 1 = present).

3.2.Exploratory Analysis of Data(EAD)

We reviewed the distribution and inter-variable correlations of the dataset extensively as part of the Exploratory Data Analysis (EDA) for our HD prediction research. A diversified population is shown by the generally normal distribution of the histograms for continuous variables such as age, resting blood pressure (trestbps), serum cholesterol (chol), and maximal heart rate reached (thalach). Bar charts for categorical variables show, as shown in Figure 2, that men are more likely to have some forms of chest pain (cp), although usual angina is less prevalent. Additionally, the charts show a range of resting electrocardiographic findings (restecg).

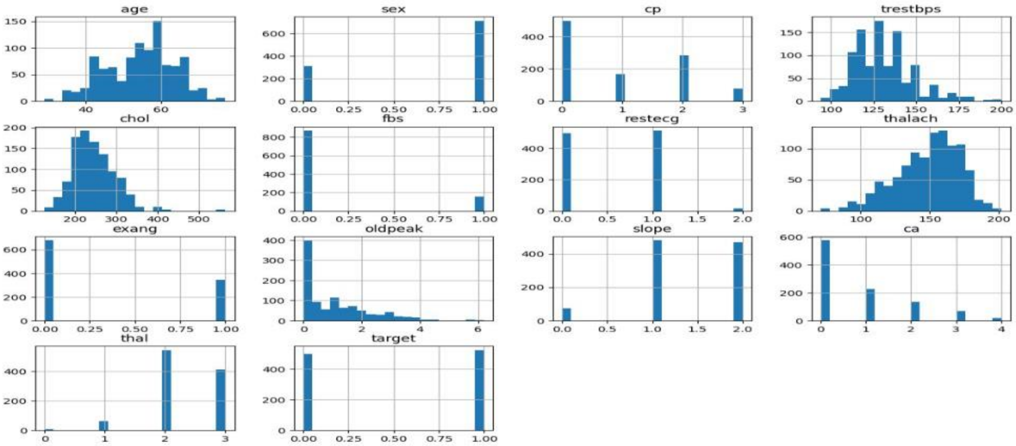


Figure (2): Clinical Attribute Distribution in Heart Disease

3.3.Pre-processing

Data pieces make up a database. You could also call them records, data sets, events, or notes. Nevertheless, each one is explained using various traits. That's what people in data science call them: characteristics or features.

Prior to creating a model with these traits, the data must be processed.

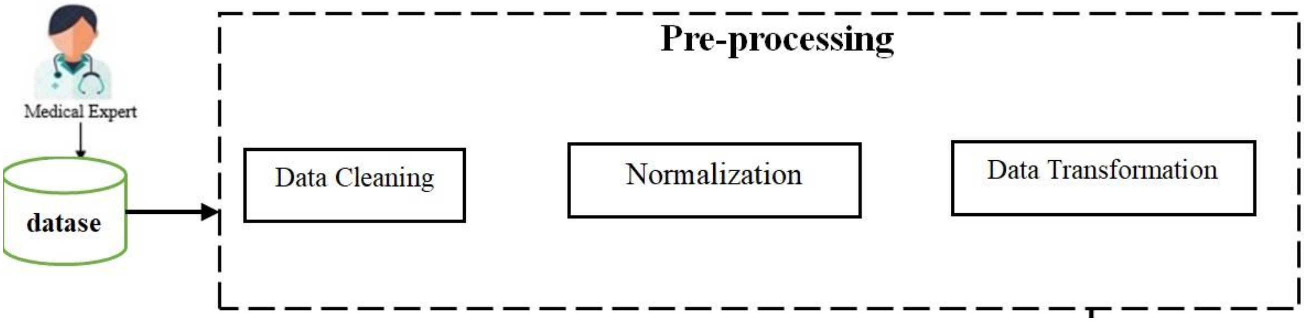


Figure (3) : Pre – processing

- **Cleaning Data**

The goal of cleaning data is to give machine learning sets of examples that are simple, full, and clearly laid out.In some cases, a data set is missing info. After finding it, we find the mode, mean, or median of the trait and fill in the numbers that are missing. This can virtually make the data set more varied.

- **Transformation Data**

Cleaning and smoothing the data is actually the same thing as changing the data. For example, we know how to change the data into a shape that the machine can understand by "transforming" it.

- Grouping

Collecting data and putting it in a standard manner for study is what data aggregation is all about.Machine learning models give more accurate results when they are fed a lot of high-quality data.

- **Normalization**

It helps keep the data within a certain range so that you don't make mistakes when training and/or analyzing the data-driven machine learning models. A broad range of data makes it hard to compare the numbers. The original data can be changed linearly, decimally scaled, or normalized using Z-scores, among other normalization methods.

Here are the methods for changing or expanding data

Data Pre-processing: This is the process by which all of the stages are framed on data that will be used in our model training procedure for HD shortly after. At this point, the dataset was divided into features (X) and target variables (y). The target, which is separated from our feature collection, indicates whether or not the person has HD. By removing the target column from our original dataframe (df), we are able to achieve a separation. This creates a new dataframe named X, which contains all of the independent variables wish to utilize for prediction. In parallel, I establish a variable (y) that will act as Ground Truth for supervised training in this scenario and store the "target"/"label" column.

This stage can only standardize this specific feature set of data because that is done after dividing the dataset into features and target. Using features with varying sizes and ranges necessitates this crucial preprocessing step, which goes by the name of normalization. In the process of normalizing or scaling each feature so that, given features with data that is roughly normally distributed, the mean is zero and the standard deviation is one. Since everything is scaled to the same level, no feature is dominating the model.

StandardScaler, a comparable tool for normalizing data, is included in the sklearn library preprocessing module. This will determine each feature's mean and standard deviation within the dataset. After learning these statistics, a scaler divides and subtracts mean and standard deviation from a dataset's features. Next, using the function fit_transform, a scaler is fitted to the feature set X, resulting in the creation of the converted dataset (X_scaled). Since each feature in this dataset is measured at a same normalized magnitude scale, they all contribute impartially to the model.

3.4. Classification Techniques

In this study used two techniques of classification for heart diseases predication as following :

3.4.1. Random forest (RF)

Random Forest solves Regression and Data Classification challenges. This strategy works well with data sets that have both continuous variables for use in regression problems and categorical ones for use in classification tasks. When dealing with classification issues, Random Forest is usually the better choice .

The Random Forest Classifier's accuracy improves as the number of trees increases . Constructing a large number of decision trees using data samples selected at random is the initial stage of the process. Then, an enhancement over using separate decision trees is made: the forecasts from all the trees are merged, and the ultimate option is made by a majority vote. Examples of possible uses include image classification, recommendation systems, and feature selection. The ensemble of many trees not only helps to avoid overfitting but also improves the algorithm's accuracy and durability. When we average the predictions produced by each tree, we can reduce biases.

In a supervised learning setting, this classifier makes use of several Decision Tree Classifiers. These trees frequently detect outliers with minimal changes to the training model, reducing feature variability in the dataset. Classification accuracy is improved using this method on the same training set. Data sets are evaluated with the caveat that a small amount of bias could be introduced during testing. To solve challenges, Random Forest employs an ensemble approach. Trees often select that one as their output class.

Incorporating the Random Forest into a prediction model requires the following steps: The dataset's data is preprocessed.

1-The dataset's data has been preprocessed.

2-Here we will make predictions about the results of the test set.

3-The training set is subjected to the Random Forest algorithm.

4-Here are the anticipated outcomes for the test set.

Accuracy is verified, and performance is usually evaluated using a confusion matrix.

6-The results of the test set are then displayed.

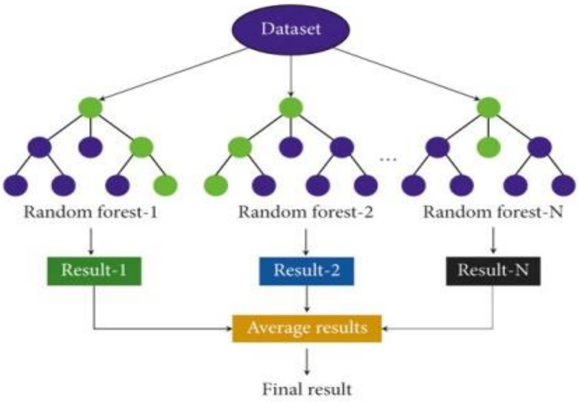


Figure (4): Architecture of (RF) [30]

3.4.2. Support Vector Machine(SVM)

One popular discriminative classifier in supervised machine learning, the SVM is used for regression and classification issues.SVM's superior prediction accuracy compared to other algorithms and its capacity to manage complex, large datasets have made it well-known . Using training datasets with labeled input-output pairs, it determines the best hyper plane for greatest margin class distinction. the number .

When traditional linear methods fail to handle a problem, support vector machines (SVMs) with kernel strategies can handle complex difficulties with non-linear connections. This method downsizes the classification problem by transforming the data into a higher-dimensional space. It divides the input values using hyper planes, taking into account just the points needed for class identification. In support vector machine (SVM) applications, finding the hyper plane with the largest margin for optimal class separation is the main goal.

Several medical jobs can benefit from SVM, including the following: estimating cardiovascular risk, detecting different cardiovascular illnesses, interpreting QRS complexes in echocardiograms, and finding aberrant heart sounds .

Iteratively creating hyper planes is the best way to split the classes:

(1) The model's SVM follows a predetermined procedure.

2–Next, select the hyper plane that divides the classes optimally; this will serve as the model's decision boundary.

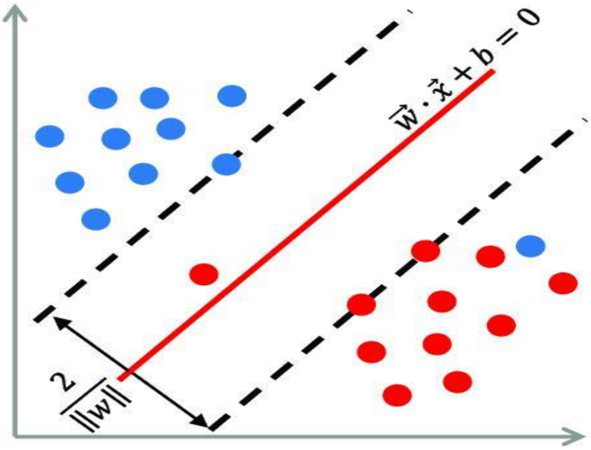


Figure (5) : An illustration of the. SVM [32].

4. System Assessment

In this stage will be explain the method was used for the proposed system evaluation . A confusion matrix classifies predictions as either true positives, false positives, true negatives, or false negatives, making it a helpful tool for evaluating model performance. This particular case involves a matrix that is specifically engineered to differentiate between the presence ('1') and absence ('0') of heart illness. What the matrix shows is

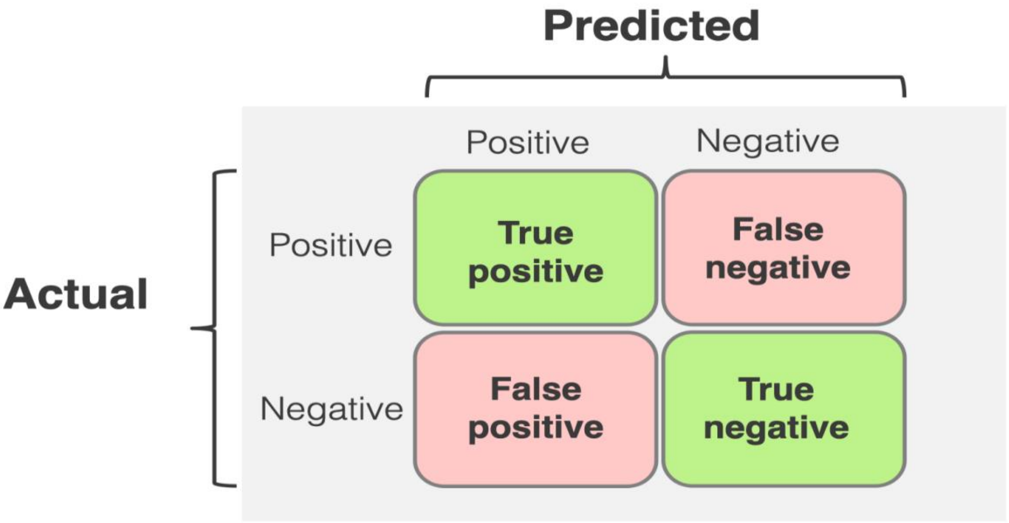
• (TP): The diagnosis of heart disease was correct in cases.

•(TN): The cases were appropriately identified as lacking cardiac disease.

•(FP): The model produced an astounding zero false positives.

•(FN): when the people were wrongly identified as not having cardiac disease.

The importance of reducing false negatives in medical diagnostics is shown by the CM in Figure 6, which measures the RF classifier's effectiveness in identifying heart sickness and also provides critical suggestions for further developing the model.



		Predicted	
		Positive	Negative
Actual	Positive	True positive	False negative
	Negative	False positive	True negative

In Random Forest algorithm's utility for cardiac illness classification, we undertake a thorough examination of its performance metrics. An ensemble of decision trees that collaborate to improve prediction abilities, Random Forest is renowned for its robustness and precision. To evaluate the model's performance in detecting cardiac disease, one must refer to table1, which shows the results of the Random Forest classifier.

The model works admirably; its 98.5% accuracy rate is evidence of its superior situational awareness. An astounding 97% sensitivity makes it ideal for detecting instances of heart disease, and a perfect 100% specificity ensures that the ailment is not present in healthy individuals.

Table2 gives a bird's-eye perspective of how well the RF classifier identified HD. When taken as a whole, these measurements demonstrate that the RF classifier is a reliable and accurate technique for detecting cardiac problems. Its competence in detecting instances of heart disease and accurately ruling out the illness when it is not present is demonstrated by its excellent recall and precision in both categories.

Table 2: Results of RF

	Precision %	Recall %	F1- Score %	Support
No Heart Attack	**0.97**	**1.00**	**0.99**	**102**
Heart Attack	**1.00**	**0.97**	**0.99**	**103**
Accuracy			**0.99**	**205**
Macro avg	**0.99**	**0.99**	**0.99**	**205**
Wighted avg	**0.99**	**0.99**	**0.99**	**205**

In order to determine how well the SVM classifier detects cardiac disease, it must first undergo assessment. In the same cases ,we can see the SVM classifier's confusion matrix, which is used to determine if heart disease is present (1) or not (0). The model's accuracy in case classification is shown by the matrix. Using the matrix as a guide, the SVM detected 97 instances of heart illness and correctly recognized 58 cases of no heart disease. Nevertheless, six instances were erroneously identified as having heart disease, and seventeen were erroneously deemed to have no heart disease at all. To assess the efficacy of the SVM in diagnosing cardiac illness, this matrix is essential as it gives a comprehensive breakdown of the classifier's performance.

- The SVM accurately recognized 97 instances of cardiac illness.
- It also properly categorized 58 people as having no heart disease.

False negatives occurred in 17 cases because the condition was wrongly classified as not having cardiac disease.

- Furthermore, six cases were mistakenly identified as having cardiac illness, suggesting false positives.

To understand how well the SVM works for detecting heart illness in general, it is necessary to look at its confusion matrix in detail. The accuracy and reliability of the model may be evaluated critically, which might impact future updates or additions to improve its diagnostic skills. When it comes to classifying instances of heart disease, these metrics offer a detailed picture of the SVM's strengths and limitations. In particular, they highlight the model's ability to detect cardiac illness in its absence and highlight areas that can benefit from further improvement to improve diagnostic accuracy. Table3 shows the efficacy of SVM classifier in detecting heart disease .

Table 3: Results of SVM

	Precision %	Recall %	F1- Score %	Support
No Heart Attack	0.93	0.83	0.88	102
Heart Attack	0.85	0.94	0.89	103
Accuracy			0.89	205
Macro avg	0.89	0.89	0.89	205
Wighted avg	0.89	0.89	0.89	205

5. Results discussion

We put the aforementioned methods to the test using a dataset that has already been pre-processed. The broad performance metrics described before can be inferred using the confusion matrix. A confusion matrix is a useful tool for describing the model's performance. in the confusion matrix that was generated for particular algorithms using the suggested paradigm. Table 3 shows the accuracy value, recall, F1-score, and precision for SVM, whereas Table 2 shows the same data for RF. Compared to the other machine learning algorithms utilized in this research, the findings demonstrate that RF has a greater level of accuracy. It outperforms the other machine learning techniques utilized in this work in terms of accuracy (98.6%), the sensitivity was(97%) and specificity was(100%), whereas the accuracy that achieved in SVM was (88.8%),%), the sensitivity was(94%) and specificity was(83%),therefore RF being more cost-effective. Table 4 explain the summary of results of classification

Table 4: performance of the compression

	Accuracy %	**Sensitivity %**	**Specificity %**
SVM	**88.78**	**94.17**	**83.33**
Random Forest	**98.54**	**97.09**	**100.0**

Figure 6 shows the results that achieved in random forest and SVM techniques of and shows RF technique that achieved results more effect than SVM technique for HD prediction.

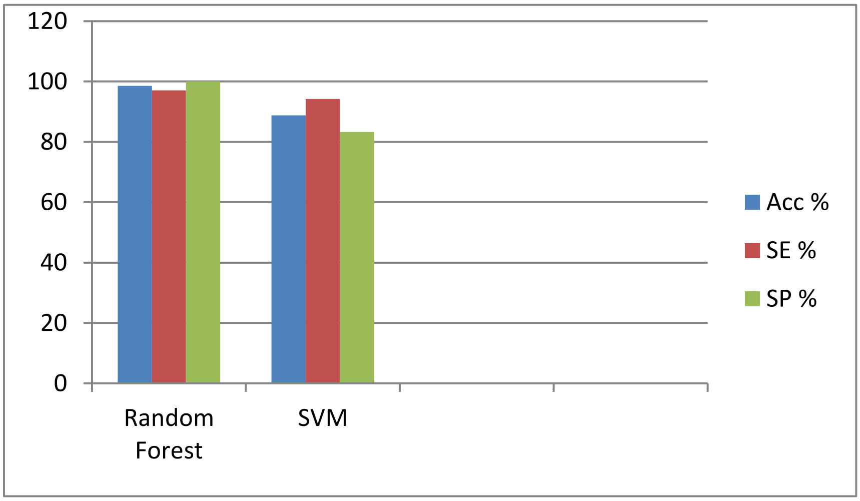


Figure 6: Comparison algorithms Performance

6. Conclusion

An precise method of predicting the onset of cardiovascular disease is urgently needed due to the rising mortality toll from this cause. The research set out to determine which machine learning parameters were most useful for detecting CHD. This research compares the performance of support vector machine and random forest in predicting the occurrence of heart disease using a dataset from a machine learning repository. With a 98.6% success rate, random forest outperforms all other green algorithms when it comes to forecasting cardiac problems. Building a radio frequency-based Internet software and employing a bigger dataset than the one utilized in this research might make future technological work more suitable for assisting fitness specialists in efficiently and successfully predicting the occurrence of heart disease.

Bibliography

[1]Seckeler, M. D., & Hoke, T. R. (2018). “The worldwide epidemiology of acute rheumatic fever and rheumatic heart disease”. Clinical epidemiology, 67-84.

[2]Saxena, K., & Sharma, R. (2016). “Efficient heart disease prediction system”. Procedia Computer Science, 85, 962-969.

[3]Dangare, C. S., & Apte, S. S. (2017). “Improved study of heart disease prediction system using data mining classification techniques”. International Journal of Computer Applications, 47(10), 44-48.

[4]Shorewala, V. (2021). “Early detection of coronary heart disease using ensemble techniques”. Informatics in Medicine Unlocked, 26, 100655.

[5]Murthy, H. N., & Meenakshi, M. (2016, November). “Dimensionality reduction using neuro-genetic approach for early prediction of coronary heart disease”. In International conference on circuits, communication, control and computing (pp. 329-332). IEEE.

[6]Al-Saedi, K. H., & Abd Alhassn, R. A. A. (2018). Miner Alerts Module to Generate Itemsets Based on FP-Growth Algorithm Improvement. Al-Mustansiriyah Journal of Science, 29(1), 114-117.

[7]Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., Blaha, M. J., Cushman, M., ... & Turner, M. B. (2016). Heart disease and stroke statistics—2016 update: a report from the American Heart Association. circulation, 133(4), e38-e360.

[8]Maiga, J., & Hungilo, G. G. (2019, October). “Comparison of machine learning models in prediction of cardiovascular disease using health record data”. In 2019 international conference on informatics, multimedia, cyber and information system (ICIMCIS) (pp. 45-48). IEEE.

[9]Soni, J., Ansari, U., Sharma, D., & Soni, S. (2019). “Predictive data mining for medical diagnosis”: An overview of heart disease prediction. International Journal of Computer Applications, 17(8), 43-48.

[10]Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M., & Qureshi, N. (2017). “Can machine-learning improve cardiovascular risk prediction using routine clinical data?”. PloS one, 12(4), e0174944.

[11]Ramalingam, V. V., Dandapath, A., & Raja, M. K. (2018). Heart disease prediction using machine learning techniques: a survey. International Journal of Engineering & Technology, 7(2.8), 684-687.

[12]Mohan, S., Thirumalai, C., & Srivastava, G. (2019). “Effective heart disease prediction using hybrid machine learning techniques”. IEEE access, 7, 81542-81554.

[13]Nabeel, M., Majeed, S., Awan, M. J., Muslih-ud-Din, H., Wasique, M., & Nasir, R. (2021). “Review on Effective Disease Prediction through Data Mining Techniques”. International Journal on Electrical Engineering & Informatics, 13(3).

[14]Ramesh, T. R., Lilhore, U. K., Poongodi, M., Simaiya, S., Kaur, A., & Hamdi, M. (2022). “Predictive analysis of heart diseases with machine learning approaches”. Malaysian Journal of Computer Science, 132-148.

[15]Alotaibi, F. S. (2019). “Implementation of machine learning model to predict heart failure disease”. International Journal of Advanced Computer Science and Applications, 10(6).

[16]Shah, D., Patel, S., & Bharti, S. K. (2020). “Heart disease prediction using machine learning techniques”. SN Computer Science, 1, 1-6.

[17]Shimaa Ouf, A. I. (2021). “A proposed paradigm for intelligent heart disease prediction system using data mining techniques”. Journal of Southwest Jiaotong University, 56(4).

[18]Drożdż, K., Nabrdalik, K., Kwiendacz, H., Hendel, M., Olejarz, A., Tomasik, A., ... & Lip, G. Y. (2022). “Risk factors for cardiovascular disease in patients with metabolic-associated fatty liver disease”: a machine learning approach. Cardiovascular Diabetology, 21(1), 240.

[19]Boukhatem, C., Youssef, H. Y., & Nassif, A. B. (2022, February). “Heart disease prediction using machine learning”. In 2022 Advances in Science and Engineering Technology International Conferences (ASET) (pp. 1-6). IEEE.

[20] Chandrasekhar, N., & Peddakrishna, S. (2023). “Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization”. Processes, 11(4), 1210.

[21]Khan, A., Qureshi, M., Daniyal, M., & Tawiah, K. (2023).”A Novel Study on Machine Learning Algorithm-Based Cardiovascular Disease Prediction”. Health & Social Care in the Community, 2023.

[22]Bizimana, P. C., Zhang, Z., Asim, M., El-Latif, A., & Ahmed, A. (2023). “An Effective Machine Learning-Based Model for an Early Heart Disease Prediction”. BioMed Research International, 2023.

[23]Beyene, C., & Kamat, P. (2018). “Survey on prediction and analysis the occurrence of heart disease using data mining techniques”. International Journal of Pure and Applied Mathematics, 118(8), 165-174.

[24]Fitriyani, N. L., Syafrudin, M., Alfian, G., & Rhee, J. (2020). “HDPM: an effective heart disease prediction model for a clinical decision support system”. IEEE Access, 8, 133034-133050.

[25] Chowdhury, M. N. R., Ahmed, E., Siddik, M. A. D., & Zaman, A. U. (2021, April). “Heart disease prognosis using machine learning classification techniques”. In 2021 6th International Conference for Convergence in Technology (I2CT) (pp. 1-6). IEEE.

[26]Kumari, A., & Mehta, A. K. (2021, August). “A novel approach for prediction of heart disease using machine learning algorithms”. In 2021 Asian Conference on Innovation in Technology (ASIANCON) (pp. 1-5). IEEE.

[27]Justo-Silva, R., Ferreira, A., & Flintsch, G. (2021). “Review on machine learning techniques for developing pavement performance prediction models”. Sustainability, 13(9), 5248.

[28]Sruthi ER. Understanding random forest. https://www.analyticsvidhya. com/blog/2021/06/understanding-random-forest/, 2022.

[29]Goel, R. (2021, July). “Heart disease prediction using various algorithms of machine learning”. In Proceedings of the International Conference on Innovative Computing & Communication (ICICC).

[30]Fu, H., & Qi, K. (2022). “Evaluation Model of Teachers' Teaching Ability Based on Improved Random Forest with Grey Relation Projection”. Scientific Programming, 2022, 1-12.

[31]Liu, G., Mao, S., & Kim, J. H. (2019). “A mature-tomato detection algorithm using machine learning and color analysis”. Sensors, 19(9), 2023.

[32]Raizada, R. D., & Lee, Y. S. (2013). “Smoothness without smoothing: why Gaussian naive Bayes is not naive for multi-subject searchlight studies”. PloS one, 8(7), e69566.

[33]Sánchez, A. S., Iglesias-Rodríguez, F. J., Fernández, P. R., & de Cos Juez, F. J. (2016). Applying the K-nearest neighbor technique to the classification of workers according to their risk of suffering musculoskeletal disorders. International Journal of Industrial Ergonomics, 52, 92-99.

[34] Kilic, A. (2020). Artificial intelligence and machine learning in cardiovascular health care. The Annals of thoracic surgery, 109(5), 1323-1329.

[35]Yousif, H., Al-saedi, K. H., & Al-Hassani, M. D. (2019). Mobile phishing websites detection and prevention using data mining techniques. iJIM, 13(10), 205.

[36]Krittanawong, C., Zhang, H., Wang, Z., Aydar, M., & Kitai, T. (2017). “Artificial intelligence in precision cardiovascular medicine”. Journal of the American College of Cardiology, 69(21), 2657-2664.

[37]Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019). “Comparing different supervised machine learning algorithms for disease prediction”. BMC medical informatics and decision making, 19(1), 1-16.

[38]Aparna, P., & Sharma, K. M. (2020, March). “Detection of A Fib and its Classification using SVM”. In 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA) (pp. 116-120). IEEE.

[39]Noble, W. S. (2006). “What is a support vector machine?”. Nature biotechnology, 24(12), 1565-1567.

[40]Goel, R. (2021, July). “Heart disease prediction using various algorithms of machine learning”. In Proceedings of the International Conference on Innovative Computing & Communication (ICICC).

[41]Fareed, M., & Al-Saedi, K. H. (2022, November). Proposal to enhance the information security system of the cloud using data mining techniques. In AIP Conference Proceedings (Vol. 2394, No. 1). AIP Publishing.