

Towards Proactive Cardiovascular Health: Machine Learning Models for Early Detection of Heart Disease

Aakansha Tyagi

KIET Group of Institutions, Uttar Pradesh

aakansha.2024cs1004@kiet.edu

Ayush Jaiswal

KIET Group of Institutions, Uttar Pradesh

ayush.2024cs1134@kiet.edu

Vivek Kumar Sharma

Department of Computer Science

vivek.sharma@kiet.edu

Aakriti Gupta

KIET Group of Institutions, Uttar Pradesh

aakriti.2024cs1181@kiet.edu

Ayush Tyagi

KIET Group of Institutions, Uttar Pradesh

ayush.2024cs1200@kiet.edu

Abstract – Cardiovascular diseases (CVDs), also known as heart disease, has replaced other diseases like malaria, strokes, cancer, trachea, bronchus which was once considered the most significant diseases worldwide, as the number one killer globally, notably this is true for India as well as other countries. This means that such illnesses should be identified early enough for effective treatment to be undertaken hence there should be such as reliable and appropriate system. Using machine learning on various medical datasets, these approaches have automated large scale and complex data analysis. For the advancement in growth of healthcare our primary goal of this study work is to determine which patient, depending on several medical parameters, is more likely to develop a cardiac condition. To determine whether the patient is likely to receive a heart disease diagnosis or not, we developed a heart disease prediction algorithm. use the patient's medical history. We employed many machine learning algorithms, including KNN and logistic regression, support vector machine, random forest algorithm are used to categorize and predict heart disease patients. Quite beneficial method was employed to control the model's application in order to increase the forecast accuracy of a heart attack in any person. The suggested model's strength was rather pleasing; it could identify signs of heart illness in a specific person using support vector machines, which demonstrated high accuracy when compared to earlier classifiers like KNN and logistic regression, etc. Thus, by utilizing the provided model to determine the likelihood that the classifier can correctly and precisely diagnose cardiac illness, a considerable amount of pressure has been released. The Given heart disease prediction system lowers costs and improves medical treatment. This research provides us with important information that may be used to forecast the people who will have heart disease.

Keywords— Cardiovascular Disease, Demographic Factors, Proactive Health, Predictive Modeling, Genetic Markers.

I.INTRODUCTION

The World Health Organization (WHO) reports that heart disease is responsible for more than 31% of fatalities

worldwide [1]. For the past 15 years, cardiovascular diseases (CVD) have been the leading cause of death. If one were to concentrate on India, the WHO estimates that heart-related illnesses cost the nation \$237 billion between 2005 and 2015. According to the European Public Health Alliance (European Public Health Alliance 2010) heart attacks, strokes, and other circulatory illnesses account for 41% of all fatalities. According to projections made by the Australian Institute of Health and Welfare (AIHW), cardiovascular disease (CVD) would account for 42% of all fatalities in Australia in 2018 [2].

Many symptoms can indicate heart disease, making a quicker and more precise diagnosis difficult. The heart is the primary organ in the human body. There are two types of heart disease risk factors. Heart issues may or may not be managed. Clinical data suggests that an individual's risk of heart disease, or more specifically, cardiovascular disease, is increased by uncontrollable circumstances. A high-fat diet, obesity, hypertension, low levels of HDL (good) cholesterol, high rates of LDL (bad) cholesterol, and a family history of cardiovascular disease are only a few of these risk factors, per Ref. Smoking, drinking, being overweight, having high blood pressure, and having high cholesterol are all risk factors that people may control. Doctors usually make diagnoses by looking at a patient's current health status, previous diagnoses made on individuals with the same status, and other factors.

As per the September 15, 2017 release of the 2016 Global Burden of Disease Report, heart diseases claimed the lives of 1.7 million Indians in 2016 [3]. In addition to raising health care costs, heart-related illnesses lower personal productivity. Globally, medical organizations gather information on a range of health-related topics. A variety of machine learning techniques can be applied to these data to obtain insightful knowledge. However, the amount of data being gathered is enormous, and it frequently contains a lot of noise. These datasets can be easily explored with a variety of machine learning techniques, even though they are too large for human minds to process. As a result, these algorithms have recently shown to be extremely helpful in accurately predicting the presence or absence of heart-related diseases. This research paper examines four machine learning algorithms: random forest, k-nearest neighbor, logistic regression, and assistive vector machine Information technology is being used more and more in the healthcare sector to support physicians in their

decision-making processes. It supports medical professionals in managing illnesses, prescribing drugs, and identifying trends and connections in diagnosis data. Many individuals who would benefit from preventive treatment are not identified by current methods of predicting cardiovascular risk, and others end up receiving needless intervention. Using machine learning to take advantage of intricate relationships between risk factors presents a chance to increase accuracy. We investigated whether machine learning could enhance the prediction of cardiovascular risk.

Today, many research workers use diverse machine learning models to help the physicians make a correct cardiac disease diagnosis. The brain has top priorities and the heart comes next in the body of individual. The blood is pumped by the heart throughout the body, feeding each organ with it. It is an important task in the medical field to predict the incidence of heart disease. By using additional data, data analytics can be applied to predict different diseases; therefore, the medical center can forecast. Some data mining and machine learning approaches that may forecast heart disease including logistic regression, K-Nearest neighbor, random forest, and support vector machine and other. Medical professionals in the United States, as well as others countries are finding it difficult to diagnose and prognosticate heart disease. As there are so many deaths associated with heart disease it is important to find an efficient and quick way of detecting this condition to reduce death rate. Our goal in this study is to find out how well different machine learning algorithms predict heart disease. To accomplish this, we used a number of strategies. The study's dataset is accessible to the general public on Kaggle. Python was used for all of the computation, preprocessing, and visualization on Google Collab Up to 94% of heart disease cases have been predicted accurately in the past utilizing machine learning techniques.

II.LITERATURE REVIEW

With the help of different DM methods, research on heart disease prediction has been ongoing over the past many years. The majority of research have employed methods like SVM (Support Vector Machine), KNN (K Nearest Neighbor), regression techniques, Decision tree, NB (Naive Bayes) classifiers, etc. across multiple patient datasets throughout across the globe. The most recent method to foresee heart disease data is reviewed here.

Mandakini Priyadarshani Behera et al. [4] provided an algorithm for heart and liver disease prediction, as well as heart and liver datasets. SVM, PSOSVM, CPSOSVM, and CCPSOSVM are four algorithms used to predict heart and liver disease. The performance of each method has been calculated and assessed in terms of the confusion matrix, classification accuracy, classification error rate, precision, recall, and F1 score. The heart disease dataset was collected from the University of California, Irvine (UCI), machine learning repository. The heart dataset has 13 attributes and 270 instances. The variable to be predicted is 1 for the absence of heart disease or 2 for the presence of heart disease, and the dataset for liver disease is retrieved from the UCI machine learning repository and the dataset has 10 attributes and 583 instances. In the experiment, the dataset is divided into two groups: training and testing. The ratios for the training set are 80% and 20%, respectively. The experiment is carried out with the Python programming language and the libraries pandas, sci-kit learn, and pyswarm. The study concludes from rigorous experimental investigation that designed CCPSOSVM gives excellent classification results, with the highest classification rate and lowest error rate for heart and liver sickness prediction.

Se Yong Jang et al. [5] after the background research on the topic of heart patient mortality rate they found that the majority of risk prediction models are limited to specific medical conditions, limiting their applicability to general medical populations so in contrast to this, the MARKER-HF (Machine learning Assessment of Risk and Early Mortality in Heart Failure) risk model was developed. This study is about the testing of MARKER-HF's ability to predict 1-year mortality in a large community-based hospital registry database of patients with and without HF. This study included 41,749 consecutive patients who underwent echocardiography in a tertiary referral hospital (4,640 with HF and 37,109 without). The result of this study shows that the average age of the 41,749 patients was 65, with 56.2% being male. The receiver operated area under the curves for MARKER-HF prediction of 1-year mortality in patients with HF were 0.729 (95% CI: 0.706-0.752) and 0.770 (95% CI: 0.760-0.780) in patients without HF. MARKER-HF accurately predicted mortality in subgroups with and without cardiovascular disease, as well as in patients with acute coronary syndrome, atrial fibrillation, chronic obstructive pulmonary disease, chronic kidney disease, diabetes mellitus, or hypertension. Patients with cancer had a higher mortality rate at a given MARKER-HF score than patients in the other groups. The study concludes that MARKER-HF predicts mortality for patients with HF as well as for patients suffering from a variety of diseases.

Alireza Ghasemieh et al. [6] explains the fact that early detection of heart complications is extremely beneficial in the treatment of patients suffering from cardiovascular disease. This study proposes a novel model for detecting emergency readmission of heart disease patients by effectively identifying patients who require emergency assistance prior to the onset of heart attacks or other heart-related complications. To improve detection performance, a robust Stacking Ensemble Learner (SEL) is built using ensemble learning. Their SEL method predicts whether a patient with heart problems needs to be admitted as an emergency case following a preliminary admission. To ensure robustness and high accuracy in prediction results across multiple runs, the XGBoost is used as a meta-learner in the SEL model. Based on the experimental results, they conclude that the proposed stacking model is a potentially more robust approach than the individual model approach for distinguishing emergency cases from non-emergency cases. They used a novel feature integration technique to distinguish data set records with "emergency" case labels from those with "non-emergency" case labels in this work. The experimental results show that the stacking model outperforms baseline models such as logistic regression, k-nearest neighbor, decision tree, Random Forest, support vector machines, bagging, and boosting in terms of accuracy, recall, and F1 score. The SEL model predicted emergency readmissions of heart disease patients with an accuracy of 88%, which is very promising for the production-ready model in clinical practice.

Yue Huang et al. [7] provides a model to compute the coronary artery calcification score (CACS) obtained from coronary artery computed tomography angiography (CCTA) examination and combine it with the coronary artery calcification (CAC) influencing factors, which is then analyzed by machine learning (ML) to predict the probability of coronary heart disease (CHD). The Agatston score was used to quantify the level of CAC. They gathered 31 variables, including hypertension, diabetes, smoking, and hyperlipidemia, among others, to compare the correlation between the CACS and clinically relevant factors. The risk of CHD based on CACS and clinically relevant factors was

evaluated using ML models that included the random forest (RF), radial basis function neural network (RBFNN), support vector machine (SVM), K-Nearest Neighbor algorithm (KNN), and kernel ridge regression (KRR). In terms of accuracy (ACC) (78.96%), sensitivity (SN) (93.86%), specificity (Spe) (51.13%), and Matthew's correlation coefficient (MCC) (0.5192), RF outperforms the other four machine learning models. The analysis of computer ML models confirmed the importance of CACS in predicting the occurrence of CHD, particularly the RF model, which has made another advance of AI computer learning methods in the field of medical analysis due to its outstanding performance of high accuracy, capacity to handle large data sets, stochasticity, better noise immunity, balancing error, and fast learning speed.

In order to forecast cardiovascular disease Chun-Ling Lin et al. [8] provides research study which was to create a smart cardiovascular measurement system that would assess three health conditions: Shanghuo syndrome, cold hands and feet, and sleep deprivation—using ECG and PPG. The suggested approaches used a variety of machine learning techniques to evaluate the features that were extracted from the physical signal. The outcomes showed greater than 82% prediction accuracy (87% for sleep deprivation using k-nearest neighbor, 82% for the Shanghuo syndrome using ensemble learning, and 83% cold hands and feet using a kernel classifier). Moreover in the context of traditional Chinese medicine (TCM), this study also discovered novel characteristics linked to sleep deprivation, cold hands and feet, and the shanghuo syndrome. The processing techniques to extract features from PPG and ECG were proposed in this paper. In particular, all the features associated with sleep deprivation, cold hands and feet, and Shanghuo syndrome were proposed by this study, along with five features (width of vascular, tired index, cold hands and feet, and Shanghuo syndrome indices). In the field of machine learning research, an accurate prediction of Shanghuo syndrome and TCM-defined cold hands and feet while taking into account pertinent physiological features is crucial. To facilitate self-health management and collaboration with medical treatment, the developed system can be seamlessly integrated with the existing tools.

Md. Mamun Ali et al. [9] in their work, they use a survival dataset to propose machine learning (ML) for risk factors analysis and HF patient survival prediction. They used five supervised machine learning techniques on the dataset which are Decision Tree (DT), Gradient Boosting (GB), Random Forest (RF), Decision Tree Regressor (DTR), and XGBoost algorithms. They evaluate the applied algorithms' results according to F-measure, accuracy, precision, recall, and log the loss value, RF offers the best accuracy at 97.78%. This study aimed to build a ML model to predict HF survival and identify the most important risk factors. It is found that RF achieved the maximum accuracy (97.78%) along with 0.97 precision, recall, F-Measure and 0.767 log loss. Examining the risk factors based on feature importance and coefficients, displays the most predictive features. The top six variables at risk for Serum creatinine (SC), age, platelets, ejection fraction (EF), SS(SS) and creatinine phosphokinase (CPK) are all factors in HF patients. Subsequent examination of these factors reveals a notable grouping of the characteristics. The most important risk factors for HF patients, according to the survival analysis, are the increase in SC, age, and SS as well as the decrease in EF. According to their findings, HF survival can be predicted more accurately with the suggested model which has the maximum accuracy of 97.78%. Their study machine learning models are helpful for patient screening with HF probability in clinical settings.

Done Stojanov et al. [10] proposed a study in which they use feature selection techniques, the main goal of this work is to improve the prediction accuracy of heart disease. A WSVM (Weight updating Support Vector Machine) algorithm based on EGA (Enhanced Genetic Algorithm) is proposed in this paper. Machine learning classifiers have been able to predict the risk of developing heart disease with the aid of the UCI heart dataset. Using the fuzzy weighted evaluation method, the EGA-FWSVM classifier performs admirably. In the end, it was discovered that improving the classifier hyperparameters and choosing the appropriate features greatly improved prediction performance. the work can be used to predict patients whose diagnosis of heart disease is made by cleaning the dataset and applying new techniques like KNN and RF. This can result in an accuracy of 83.54% and 87.25% on average for the proposed EGA-FWSVM model, which is higher than earlier generations' accuracy of 91.68%. The idea behind a hybrid model is to combine several well-known methods for classification and selection into an ensemble model to produce impressive outcomes.

In order to predict heart diseases, Subasish Mohapatra et al. [11] have developed an efficient model in this work that combines classifier stacking and data pre-processing with outlier detection. To guarantee that the distribution of the data is uniform and on a similar scale, the data was first normalized. This guarantees the model's training stability and improves performance. Ten distinct classifiers, each with varying strengths, were used in this instance. These included instance-based (KNN, for example), probabilistic (NB, for example), and a few ensemble (XGB, GB) classifiers. They stacked different classifiers to benefit from their distinct strengths, taking into account the different prediction methods. About 92% accuracy is achieved by using MLP as the meta-learned model. With a precision of 92.6%, sensitivity of 92.6%, and specificity of 91%, the suggested stacked classifier fared better in terms of overall parameter comparison than the conventional machine learning classifiers. The suggested methodology demonstrates the benefits of merging poor learners and leveraging their heterogeneity to improve overall prediction outcomes. If CVDs are identified early on, early therapy may be able to prevent many deaths. Cardiologists cannot, however, manually evaluate the vast quantity of data gathered for a patient in order to create a timely treatment plan. Therefore, using machine learning-based systems for primary screening is a viable approach. These systems must be dependable and effective in order to use predictive analysis to diagnose patients. Comparing the model to more conventional methods, this shown high sensitivity suggests that the model has fewer false-negative outcomes.

In the US, the ACC/AHA Pooled Cohort Equations (PCE) Risk Calculator is frequently used for primary prevention of atherosclerotic cardiovascular disease (ASCVD); nevertheless, in certain groups, it may overestimate or underestimate risk. Therefore, Qi Li et al. [12] used machine learning (ML) techniques and electronic medical record (EMR) data to create an automated, population-specific ASCVD risk calculator. They then compared its prediction power to that of the PCE calculator. They gathered information from 101,110 distinct EMRs belonging to active patients. Patient datasets with either simply cross-sectional (CS) characteristics or CS mixed with longitudinal (LT) information generated from laboratory results and vital statistics were subjected to machine learning algorithms. They used a recently proposed new cost metric to compare the models usefulness. Every evaluated machine learning model outperformed the PCE risk calculator in terms of predictive ability. The combination of CS and LT features

(RF-LTC) using the random forest machine learning (ML) approach yielded the best area under the curve (AUC) score of 0.902 (95% confidence interval (CI), 0.895–0.910). The top machine learning model only needed to screen 43% of patients in order to detect 90% of all positive ASCVD cases, whereas the PCE risk calculator needed to screen 69% of patients. When compared to the PCE calculator alone, prediction models developed with machine learning approaches yielded better ASCVD prediction and required fewer screens to predict ASCVD. When LT and CS characteristics were combined in the ML models, ASCVD prediction was much better than when CS features were used only.

Khandaker Mohammad Mohi Uddin et al. [13] in their research study examines heart disease and claims that heart disease is one of the most serious ailments, killing the majority of its victims. Heart disease is very difficult to diagnose medically. This diagnosis is a difficult procedure that calls for rapidity and precision. Early identification of heart disease will reduce the risk of mortality. Predicting heart illness has grown to be one of the most challenging medical tasks in recent years due to the prevalence of cardiac issues. Investigating closely related features, researchers found the most accurate predictors of various disorders. In this work, the existence of cardiac anomalies is detected using Machine Learning (ML) approaches. Several machine learning (ML) algorithm techniques, including Decision Tree (DT), Ada-Boost Classifier (AB), Extra Trees Classifier (ET), Support Vector Machine (SVM), Gradient boost, MLP, extreme gradient boost (XGB), Random Forest (RF), KNN, and LR, are used in the proposed method to predict the likelihood of heart disease and classify the risk level of the patient. The suggested method is trained and tested using three separate datasets. The testing findings demonstrate that the Decision Tree method has the best accuracy, at 99.16%, when compared to other machine learning techniques. If features are handled effectively, success within the categories of heart disease diagnosis should be evident. Cardiovascular disease classification can be effective if certain traits are preserved

The prior research, which employed a small number of effective machine learning algorithms to build heart disease prediction models, served as the paradigm for this investigation. This suggested approach differs from the others in that it makes use of an integrated heart-disease dataset created from conventional datasets in addition to feature selection approaches. Moreover, a number of prediction models are developed and their effectiveness is evaluated by comparing them with respect to certain performance metrics such as accuracy, precision, and recall in order to determine which are the most successful in terms of potential application in the medical field for the prediction of cardiac illness.

III.PROPOSED METHODOLOGY

This section provides a detailed description of the cardiac datasets and the suggested strategy for predicting heart disease. The primary goal of the suggested approach is to forecast the onset of cardiac disease in order to quickly and accurately recognize the condition. To predict heart disease based on some health characteristics, our method uses a variety of data mining approaches and machine learning algorithms, including Support vector machines, K Nearest Neighbor (KNN), Logistic regression, and Random Forest and Google collab is used to evaluate the data as shown in flow chart fig 1. By importing libraries, we are using this open-source software to implement a

variety of machine learning techniques.

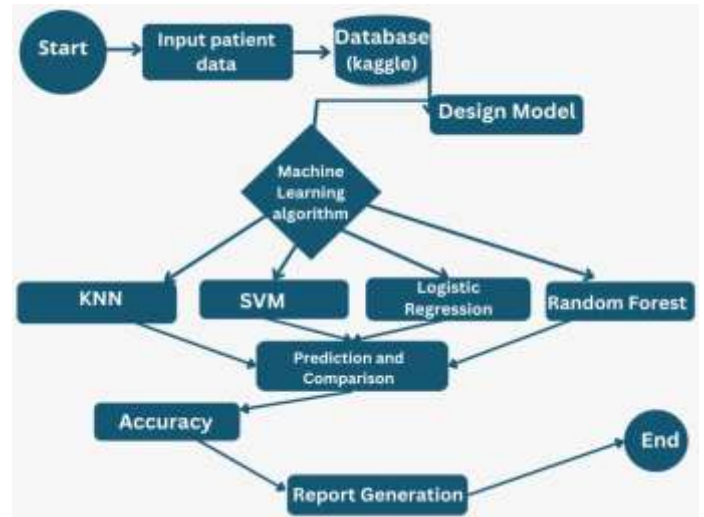


Fig 1-Flow Chart

[3.1] Dataset for implementation

We have used built in dataset from Kaggle for predicting heart disease. This database contains a total of fourteen attributes which are listed below in Table 1:

Table 1-Parameter Table

S.No:	Parameter	Information
1	age	Patient's age in years
2	sex	0 = female,1=male
3	cp	Chest pain experienced. 0 = typical angina 1 = atypical angina 2 = non-anginal pain 3 = asymptomatic
4	trestbps	Resting blood pressure (in Mm Hg on admission to the hospital)
5	chol	Serum cholesterol in mg/dl
6	fbs	Fasting blood sugar >120 mg/dl 1=true,0=false
7	restEcg	Resting electrocardiographic measurement 0 = normal, 1 = having ST-T wave abnormality,2 = showing probable or Definite left ventricular hypertrophy by Estes' criteria.
8	thalch	Maximum heart rate achieved
9	exang	Exercise-induced angina 0=false,1=true
10	oldpeak	ST depression induced by exercise relative to rest
11	slope	The slope of the peak exercise ST segment
12	ca	The number of major vessels

		(0-3)
13	thal	A blood disorder called. Thalassaemia, 3=normal
14	target	Has heart disease or not? 0 = no,1=risk

[3.2] Data splitting

Training and testing data are divided up. 20% of the data are used for testing, while 80% are used for training as shown in fig 2.

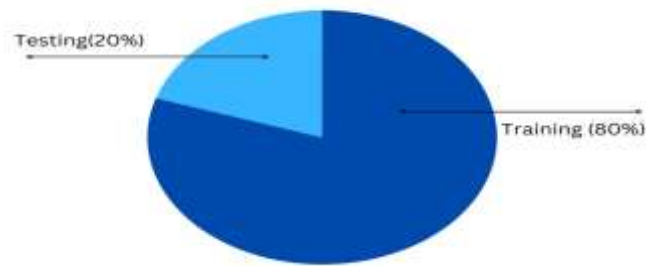


Fig 2 -Training and testing percentage

[3.3] Random Forest Algorithm

The random forest algorithm belongs to a category of supervised classification technique that consists of multiple decision trees working together as a group. The class with the most votes become the prediction made by our model. Each tree in the random forest makes a class prediction, which eliminates the limitations of the decision tree algorithm. This improves accuracy and reduces overfitting of the dataset. When used on large datasets, the random forest approach may still provide the same results even if a significant portion of record values are missing. The samples produced by the decision tree may be saved and used with various data types. In the research, random forest achieved a test accuracy of 77% as described in classification report in table 2.

Methods used for Random Forest Algorithm

```
RF_model= RandomForestClassifier(n_estimators=20)
RF_model.fit(x_train_scaler, y_train)
y_pred_RF= RF_model.predict(x_test_scaler)
RF_model.score(x_test_scaler,y_test)
```

Table 2-Classification Report

	Precision	Recall	F1-score	Support
0	0.74	0.84	0.79	38
1	0.81	0.70	0.75	37
accuracy			0.77	75
macro avg	0.78	0.77	0.77	75
weighted avg	0.78	0.77	0.77	75

[3.4] K-Nearest Neighbor

By comparing a given test characteristic with training attributes that are similar to it, K-Nearest Neighbor is a type of lazy learner classifier that relies on learning by similarity. Pattern recognition and statistical estimation are the two main applications for K-NN, which are both regarded as non-parametric techniques. Each attribute is represented by an n-dimensional space if there are n attributes for training, according to Jiawei et al. (data mining concepts, second edition book). The k-nearest neighbor classifier looks for the k-training attributes that are near to the unknown attribute in the pattern space when

it is given an unknown attribute, such as a test attribute. indicating that the closest neighbors of the unknown attribute are the k training attributes. Distance metrics, such as the Euclidean distance, are used to measure how close two attributes are to each other. In the research, K-Nearest Neighbor achieved a test accuracy of 75% as described in classification report in table 3.

Method used for K-Nearest Neighbor

```
Knn_model= KNeighborsClassifier(n_neighbors = 5, metric =
'minkowski', p = 2)
knn_model.fit(x_train_scaler, y_train)
y_pred_knn= Knn_model.predict(x_test_scaler)
knn_model.score(x_test_scaler,y_test)
```

Table 3-Classification Report

	Precision	Recall	F1-score	Support
0	0.74	0.76	0.75	38
1	0.75	0.73	0.74	37
Accuracy			0.75	75
Macro avg	0.75	0.75	0.75	75
Weighted avg	0.75	0.75	0.75	75

[3.5] Support Vector Machine

Support vectors are those that are both most distant from both sides of the hyperplane and located in close proximity to it There are two methods for multiplying vectors: dot product and cross product. The only distinction is that the cross-product output is a vector while the dot product output is a scalar value as a result.

The projection of one vector onto another multiplied by the product of another vector is known as the dot product.

In the research, Support vector machine achieved a test accuracy of 79% as described in classification report in table 4.

Method for Support vector Machine

```
SVC_model= SVC ( )
SVC_model.fit(x_train_scaler, y_train)
y_pred_SVC= SVC_model.predict(x_test_scaler)
SVC_model.score(x_test_scaler,y_test)
```

Table 4-Classification Report

	Precision	Recall	F1-score	Support
0	0.79	0.79	0.79	38
1	0.78	0.78	0.78	37
Accuracy			0.79	75
Macro avg	0.79	0.79	0.79	75
Weighted avg	0.79	0.79	0.79	75

[3.6] Logistic Regression

Logistic Regression is a type of supervised learning which uses the observations done in the past from a dataset to give the desired data. Logistic regression is one of the most important techniques used in the field of machine learning. Predictions of the classes in the datasets will improve with the number of relevant data entered in the dataset. Extract, Transform, and Load (ETL) is a method for preparing data for analysis. A logistic regression model takes into factor one or more independent factors and analyzes the relationship between them to predict an output variable. A logistic regression could be used to predict weather for answering the questions like whether it would rain tomorrow or not. The function of logistic regression is given by $p(x)$ where $p(x) = (1 / (1 + \exp(f(x))))$. As a result, it's frequently between 0 and 1. If the probability for a given x equals 1 then it is read as function of $p(x)$. As a result, the probability of output being 0 is $1 - p(x)$. In the research, Logistic regression achieved a accuracy of 76% as described in classification report in table 5.

Method for Logistic Regression

```
LR_model= LogisticRegression()  
LR_model.fit(x_train_scaler, y_train)  
y_pred_LR= LR_model.predict(x_test_scaler)  
LR_model.score(x_test_scaler,y_test)
```

Table 5-Classification Report

	Precision	Recall	F1-score	Support
0	0.753	0.79	0.77	38
1	0.77	0.7	0.75	37
Accuracy			0.76	75
Macro avg	0.76	0.76	0.76	75
Weighted avg	0.76	0.76	0.76	75

IV. RESULT AND DISCUSSION

This section includes several experiments to demonstrate how effectively the proposed model predicts cardiac disease. The suggested Support Vector Machine (SVM) model performs better than existing techniques such as logistic regression, KNN, and Random Forest (RF) as shown in graph fig 3. A range of performance metrics were employed to evaluate the efficiency of the ML models that were previously discussed.

The suggested learning model's total prediction performance is shown by accuracy. True positive and true negative refer to the categorization models' capacity to predict if a patient will acquire heart disease and if they won't. The terms false positive (FP) and false negative (FN) represent the amount of incorrect predictions made by the models. Precision and recall are used to quantify the sensitivity and precision of the heart disease classification model, respectively.

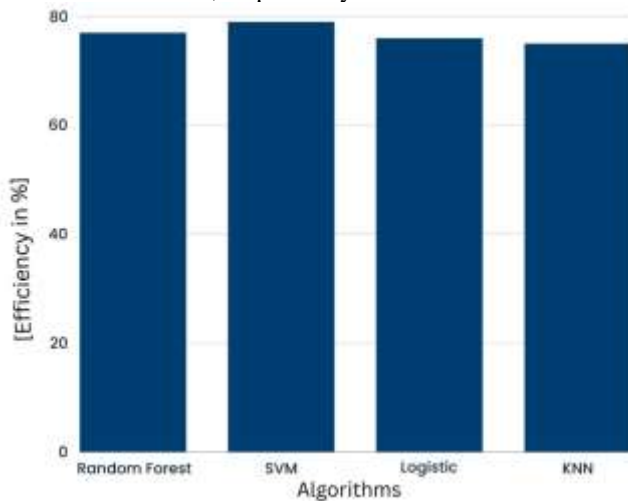


Fig 3-Efficiency vs Algorithms

The ratio of precisely measured positive observations to the total number of predicted positive observations is the definition of precision.

$$TP/TP + FP = \text{Precision}$$

Sensitivity may be calculated by dividing the total number of observations in the relevant class by the number of positive observations that were correctly identified.

$$\text{Recall} = TP/TP + FN$$

The calculation of the F1 score involves balancing accuracy and recall. It therefore takes into account both false positives and false negatives.

$$F1 \text{ Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

The following formulae are used to calculate accuracy for both positive and negative values.

$$(TP + FP)/(TP + TN + FP + FN) = \text{Accuracy}$$

This formula uses the phrases True Positive, False Positive, True

Negative, and False Negative, correspondingly.

The outcomes show that the proposed approach may accurately predict heart disease when the classifier model has access to sufficient data. These results indicate that, based on its greater performance, the proposed model outperforms earlier models by a significant margin.

V. CONCLUSION AND FUTURE WORKING

Cardiovascular diseases (CVDs) which also called heart diseases are the most life threatening disease nowadays which develops tragical compilations such as heart attack. In contrast to which we aimed to build a Machine learning system to predict cardiovascular disease. We have studied the data sets and throughout the process of developing the cardiovascular disease prediction system, various machine learning algorithms were explored and estimated such as Random Forest, K nearest neighbour, logistic regression and support vector machine, leading to the selection of the most suitable models based on their accuracy, reliability, and specificity. By implementing different ML techniques we found that support vector gives the most accurate result fig 3.

The progressive implementation of a heart disease prediction system using Machine learning techniques shows crucial advancements in field of science and health care. This study provides new ideas for the assessment method of CVD, promoting ML to have a higher prospect of application in medical fields such as predicting the risk of the disease and diagnosing the treatment. By adopting most accurate algorithms and broad datasets, ML systems can provide fast predictions about an individual's risk of developing heart diseases and helps in decreasing deaths due to cardiovascular disease. The achievable benefits of such a system are enormous, including fast detection, timely interventions, and ultimately, saving lives.

In future studies, we will collect more samples, data sets and try to add more parameters to find out the prediction fastly and accurately we also try to optimize models, and build deep learning-based models to more accurately predict the development of CVD, as well as treatment and prevention recommendations.

VI. REFERENCES

- [1] S. G. Sugendran, "Earlier identification of heart disease using enhanced genetic algorithm and fuzzy weight based support vector machine algorithm," 2023.
- [2] "Heart, stroke and vascular disease: Australian facts," 2023.
- [3] S. Yadavar, "Indiaspend," 2017. [Online]. Available: <https://www.indiaspend.com/heart-disease-kills-1-7-million-indians-continues-to-be-top-killer-in-2016/>.
- [4] A. S. D. M. ., S. Mandakini Priyadarshani Behera, "A Hybrid Machine Learning algorithm for Heart and Liver Disease Prediction Using Modified Particle Swarm Optimization with," *Procedia Computer Science*, 2022.
- [5] M. SeYongJang, "Machine Learning Model Prediction of Mortality in Patients With and Without Heart Failure," *JACC: Advances*, vol. 2, 2023.
- [6] A. L. P. B. P. V. R. K. Alireza Ghasemieh, "A Novel Machine Learning Model with Stacking Ensemble Learner for Predicting Emergency Readmission of Heart-Disease Patients," *Decision Analytics Journal*, 2023.
- [7] 1. ., Y. R. a. ., H. Y. a. ., Y. D. b. ., Y. L. a. ., Y. Y. a. .,

- A. M. a. ., T. Y. d. ., Y. W. c. ., F. X. c. ., Q. H. e. Y. Z. a. Yue Huang a, "Using a machine learning-based risk prediction model to analyze the coronary artery calcification score and predict coronary heart disease and risk assessment," *Computers in Biology and Medicine*, 2022.
- [8] C.-k. T. b. C.-J. W. b. S.-H. C. b. Y.-S. H. b. L.-J. K. b. Chun-Ling Lin a, "Title-Development of smart cardiovascular measurement system using feature selection and machine learning models for prediction of sleep deprivation, cold hands and feet, and Shanghuo syndrome," *International Measurement Confederation* , 2023.
- [9] V. S. A.-D. b. ., N. M. c. ., A. A. H. d. ., I. M. a. ., S. A. e. ., K. F. A.-t. f. ., K. A. g. ., F. M. B. g. ., M. A. M. Md. Mamun Ali a, "Machine Learning Approach for Risk Factors Analysis and Survival Prediction of Heart Failure Patients," *Healthcare Analytics*, 2023.
- [10] L. V. R. G. Stojanov, ": Machine Learning-Based Predictive Models for Heart Disease Diagnosis," *King Saud University*, 2023.
- [11] S. M. P. K. P. S. M. Subasish Mohapatra, "Heart Diseases Prediction based on Stacking Classifiers Model," *International Conference on Machine Learning and Data Engineering*, 2023.
- [12] A. C. A. R. W. E. E. Qi Li, "Automating Cardiovascular Disease Prediction Using Machine Learning and EMR Data," *International Journal of Medical Informatics*, 2022.
- [13] R. R. N. Y. N. B. S. K. D. Khandaker Mohammad Mohi Uddin, "Machine Learning-Based Approach to the Diagnosis of Cardiovascular Disease Using a Combined Dataset," *Intelligence-Based Medicine*, 2023.