

# PROGNOSIS OF CARDIOVASCULAR DISEASE USING PRINCIPAL COMPONENT ANALYSIS AND SUPPORT VECTOR MACHINE CLASSIFICATION ALGORITHMS IN THE R STUDIO ENVIRONMENT

RAHBRE ISLAM<sup>1</sup>, SAFDAR TANWEER<sup>1</sup>, TABREZ NAFIS<sup>1</sup>, IMRAN HUSSAIN<sup>1</sup>, SHAHAZAD NIWAZI QURASHI<sup>2\*</sup>

<sup>1</sup>Department of Computer Science & Engineering, School of Engineering Sciences & Technology, Jamia Hamdard University, New Delhi, India

<sup>2</sup>Department of Health Informatics, College of Public Health, and Tropical Medicine, Jazan University, Jazan, Kingdom of Saudi Arabia

<sup>1</sup>rahberislam@gmail.com, safdartaanweer@yahoo.com, tabrez.nafis@gmail.com, hussain.imran@gmail.com

<sup>2</sup>sQurashi@jazanu.edu.sa (Corresponding Author)

## ABSTRACT

Cardiovascular disease is a significant cause of death throughout the globe. Early detection of this lethal disease can be important to avoid future losses. The underlined issues can be unraveled using patients' medical history and machine learning algorithms and can predict heart disease status before it gets in worse condition. The predictive ability of ML algorithms, particularly SVM, is promising for cardiovascular diseases. This study also presents machine learning approaches for predicting heart diseases, using data on major health factors from patients. The Principal Component Analysis (PCA) and Support Vector Machine (SVM) have been applied to comprehend and classify patient data. The main aim of this study is to predict heart-related conditions well in advance to avoid any fatality. Complex data can be simplified using PCA while Support Vector Machine helps to assess predictions. The combination of these methods is applied in the R studio environment to assess heart health accurately and efficiently. Data preprocessing and feature selection steps were done before building the models. The accuracy of SVM with and without Principal Component Analysis (PCA) is 90.49% and 84.88% respectively where SVM with PCA outperformed.

**Keywords:** *Heart Disease, SVM, PCA, R Language, Machine Learning*

## 1. INTRODUCTION

Diagnosis and prognosis of any disease are important in the healthcare sector though it is considered a hectic task. It needs to be done intelligently so that its automation may be beneficial for stakeholders [1]. Sometimes clinical experts like physicians, pathologists, and even a pool of experts are unable to predict a disease. However, computer-based information systems play a vital role in reducing clinical expenses and enhancing the quality of medical care. To ensure that computer systems are working well, we should test different techniques.

In most cases, hospitals use these systems for patients and their data, but sometimes they produce lots of irrelevant information that is not very useful for making decisions in the healthcare sector. Recently, CVDs have been considered the most common illness, the world is experiencing nowadays [2]. World Health Organization (WHO) reports that more than 17.9 million heart disease casualties took place as per the fact sheet of 2019 which is an estimation of 32% of global deaths [3]. Many organizations use data mining (DM) techniques in the medical field for decision-making and identifying patterns of complex datasets [4].

Currently, many scientists are using a data mining approach to identify "how heart disease evolves". They also collected vital features for a

doctor to make better decisions. Most of them used a proven method to predict and detect heart disease in the earliest stages. They have talked about heart disease prognosis, so researchers created Machine Learning (ML) tools to help diagnose it quickly and improve treatments [5]. Only one classifier, Support Vector Machine an ML technique has been applied with and without Principal Component Analysis (PCA) to compare the outcomes of other classifiers [20].

Considering the global burden of heart disease, it is crucial to predict this disease so that life-saving measures can be taken. The underlined issues of cardiovascular diseases can be unraveled using patients' medical history and machine learning algorithms and can predict heart disease status before it gets in worse condition. This study aims to identify and utilize an effective model for early prediction of heart disease using advanced algorithms and an R Studio environment. The SVM with and without PCA will used to evaluate the assessment of cardiovascular disease accurately.

Section 2 outlines about the classification techniques and the literature based on implementing R Studio environments, Section 3 describes about the cardiovascular disease dataset and study method and required material details and Section 4 discusses the approach to solving the present problem and the importance of the machine learning classifier (SVM) to solve the problem, Section 5 Provides the result of the evaluation parameters. Section 6 represents the conclusion of the study and its prospects.

## 2. RELATED WORK

Thummala *et al.* [2023] employed logistic regression (LR) and random forest (RF) classifiers to predict heart disease and found RF with 87.64% mean accuracy and outperformed LR with a difference of 7.64% [5]. Moreover, Ziasabounchi *et al.* [2014] applied K-means, Fuzzy C-means Clustering Algorithm, with and without PCA on the same dataset Cleveland from UCI repository and K-means showed 81.0% accuracy without PCA and 87.0% accuracy with PCA [6]. Recently, Boukhatem *et al.* [2022] used multilayer perceptron, support vector machine (SVM), random forest, and Naïve Bayes, to build a

prediction model and found that SVM showed an accuracy of 91.67% among the ML algorithms [7].

Dun *et al.* [2016] studied and observed the presence of CVD by applying deep learning (DL), random forests, logistic regression, and SVM. They found NN as the best classifier with an accuracy of 78.3% among all [8]. Whereas Singh and his colleagues [2018] made heart problems easier to understand by using a ranking method called Fisher ranking, along with generalized discriminant analysis (GDA) and a binary classifier. After employing the said techniques, they improved the accuracy by 100% [9]. Yaghouby and his team also worked on arrhythmias with heart rate variability using generalized discriminant analysis (GDA) and a neural network. With this, they got a perfect score of 100% [10].

Zhang *et al.* [2018] used an Adaptive Boosting algorithm based on Principle Component Analysis (PCA) to detect breast cancer [11]. Santhanam *et al.* [2013] applied the same concept of PCA to the UCI dataset and found yielded components such as PCA1, PCA2, PCA3, and PCA4. Among these PCs, PCA1 was identified as a promising method, achieving an impressive 92.0% accuracy in regression analysis and an accuracy of 95.2% with a feed-forward neural network classifier [12]. Recently, Dhankhar & Jain [2021] explored the most accurate way to predict heart disease (HD) using a bunch of different methods from the UCI Repository. The dataset was split into 80% for training and applied different algorithms like KNN, RF, DT, and SVM. Out of all these methods, they found that the Random Forest (RF) algorithm was the best and predicted 90% accuracy [13].

Li *et al.*, [2020] created a machine-learning model that can speculate whether a person has heart problems or not. For the same, they used sophisticated math and computer tricks like KNN, DT, ANN, NB, LR, and SVM, and they also employed some special math to pick the most relevant features and observed SVM method was the best, with an accuracy of 92.37% [14]. Khan *et al.* [2017] studied and compared some well-known ML algorithms to predict CVD. They applied ANN, SVM, DT, and "repeated incremental pruning to produce error reduction" classifiers on the Cleveland dataset of the UCI repository, with 303 cases and 14 features. They found SVM with the best accuracy of 90.0% [15].

Lately, Bhatt *et al.* [2023] used ML algorithms DT, XGBoost, RF, and MLP on a real-dataset of 70,000 instances from Kaggle and found RF with an accuracy of 87.05% applying with cross-validation (CV) and 86.92% without CV [16]. Hariharan *et al.* [2018] compared different classifiers such as KNN, DT, and SVM on the VA Long-Beach dataset from the same UCI repository with 270 instances and 12 features. After evaluating the confusion matrix of the model, they concluded that SVM outperformed with an accuracy of 92.0% with 83% specificity and 100% sensitivity [17].

Gárate-Escemila *et al.* [2020] tested various methods using the combination of chi-square feature selection and PCA on Cleveland, Hungarian, and Cleveland-Hungarian datasets and found the accuracy rates were 98.7%, 99.0%, and 99.4%. The selected features from the ChiSqSelector method included relevant factors like cholesterol levels, heart rate, chest pain presence, ST depression-related features, and heart vessel information [18]. Gárate-Escamila *et al.* [2020] showed that using chi-square and PCA together improved the accuracy of most classifiers. Applying PCA to the raw data alone gave worse results and needed more dimensions to get better [19].

In the literature section, the use of feature selection strategies gives a glimpse of sophisticated model building using ML algorithms, so the prediction of heart disease may be incorporated by applying the same concept to different data sets. Hence, using other ML algorithms such as SVM with and without PCA features have been applied to the Rohilkhand Hospital dataset for heart disease prediction.

### 3. MATERIAL AND METHOD

#### 3.1 Dataset

In this study, we have obtained cardiovascular disease-related datasets from Rohilkhand Hospital, Shahjahanpur, Uttar Pradesh, India. In the initial stage, data cleaning and feature selection were done to exclude unnecessary features like noisy data, and missing values, and then classification techniques were applied using principal component analysis (PCA) and support vector machine (SVM). Here the class and structure of the data set is shown with 28 features such as 'age', 'sex', cholesterol', etc. (Figure-1) and having a total of 820 values where the Target variable is factorial. The study demonstrated that SVM with PCA performs well and its outcome unveiled a new path for diagnosing and preventing heart disease in the future.

#### 3.2 Data Preprocessing

Before data analysis, preprocessing is required for the collected data. After the treatment of missing values and transformation of numerical to categorical values of some variables of the Bundelkhand Hospital dataset, the depicted structure is shown below. This dataset contains 28 pieces of information about people, like their age, gender, blood pressure (H/L), and whether they have diabetes. We use this information to figure out if a person has heart disease or not. If they don't have heart disease, we label it as '0,' and if they do, we label it as '1.' We have used the "str()" function to look at the dataset's characteristics and then explored it further based on what we see.

Age	Ht.m2	Wt.kg	BMI	SBP	DBP	HR	PP	RBP	chol	MHR	OPK	CPT	FBS	RES	EX	slope	VCA	THA	Physical_Act	Smoking	Alcohol Drinking	HTN	Family History.of CVD	Stress	Sex	Diabeties	Target	
int	num	num	num	int	int	int	int	int	int	int	num	in	int	in	int	int	int	int	int	int	int	int	int	int	int	factorial	int	factorial

Figure 1: Structure Of Dataset

### 3.3 About R Studio Environment

R is a programming language and environment that is primarily used for statistical computing, data analysis, and data visualization. RStudio allows us to write, run, debug, and visualize R code in a user-friendly interface. A text editor with syntax highlighting, code completion, and other tools to help us write and execute R code. A window where we can interact with R directly, enter commands, and see the results.

A panel that shows the variables and objects that we have created or loaded in our R session, and their values and attributes. A panel that displays the graphical output of our R code, such as charts, graphs, and images.

Moreover, R has become a popular choice among statisticians, data scientists, researchers, and analysts for its powerful capabilities in handling data and conducting statistical analyses. Since it has a pool of useful statistical packages, which are used to manipulate and get outcomes after employing ML techniques. Some statistical packages like pROC(), ROCR(), and stats() are applied to perform and analyze the ML model.

### 3.4 Glimpse of Target Variables

Data balancing is essential for accurate results before and after the classifier is applied. The below graph shows whether the target classes are equal or not, where “1” represents heart disease patients and “0” represents no heart disease patients.



Figure 2: Number Of Patients With Or Without Heart Disease

### 3.5 Methodology, Approach, and Solution

This research study investigates an intelligent approach to predict CVD that involves using a reduced set of vital features and employs Support Vector Machines (SVM) with the combination of Principal Component Analysis (PCA) and compares their performances. Where PCA initially is used to extract relevant features before applying classifiers for prediction.

### 3.6 Principal Components Analysis (PCA)

A magical dimensionality reduction tool [20] that helps to understand the data in better ways. It evaluates a lot of information and makes it as simple as possible and keeps the important parts intact. It also shows how different groups of things are similar or not. For the same it first organizes the data and then does some calculations to see the most important things such as applying covariance or correlation matrix, later eigenvalue decomposition is applied to the correlation matrix [21].

This technique carries most of the original data in the first component, while the remaining data is explained by subsequent principal components in descending order. In other words, it is said that the first PC shows the best amount of variance [22]. To apply PCA to any dataset the central tendency and its features are considered. Here the average correlation among 12 variables of our dataset is 0.16 while with all 27 variables is 0.05. Hence the variables within this data set are eligible for principal component analysis (PCA). Some graphical representations are shown below to depict the relations among variables of the dataset.

### 3.7 Scatter Plots

This plot utilizes Cartesian coordinates to demonstrate the relationship between two variables in the dataset. The X and Y coordinates represent the values of the variables, and the data is represented as a collection of points on the plot [23].



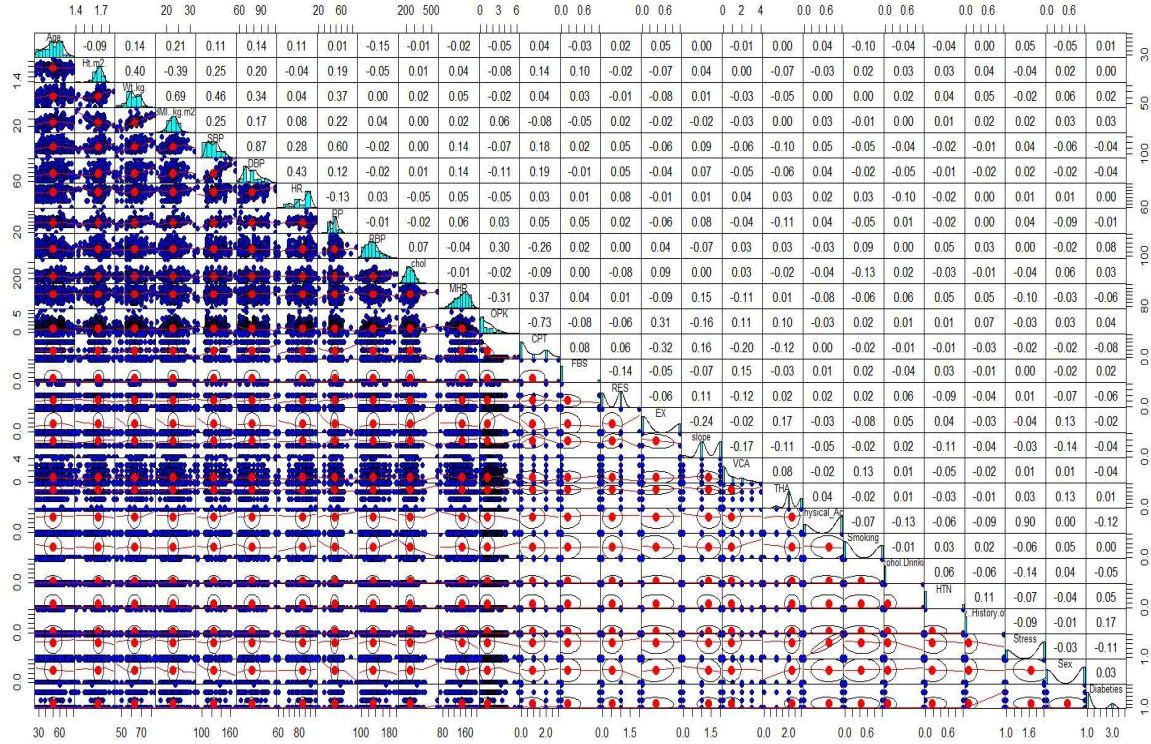


Figure 3: Scatter plot of all the variables

### 3.8 Orthogonal Feature

To reduce the dimensionality of high-dimensional data, the Principal Component Analysis (PCA) method is used. It is achieved by creating new features, known as principal components, which are linear combinations of the original features. These principal components are designed to be orthogonal to each other, capturing distinct patterns and variations within the data. Essentially, PCA helps uncover the most important aspects of the data while simplifying its representation. When the correlations among independent variables are high, that means multicollinearity rises. So, the estimates of the model may be unstable or predictions are not going to be accurate.

To handle this type of issue, Principal Component Analysis finds a way. Below the orthogonal features of independent variables having correlation coefficients '0' it means there is no dependency among independent variables

Now based on the above features and graphical representation insights of PC were applied with SVM to predict heart disease. As per the features of PC only non-categorical variables were taken into consideration, here 12 attributes carry non-categorical values which are a normalized linear combination of original variables.

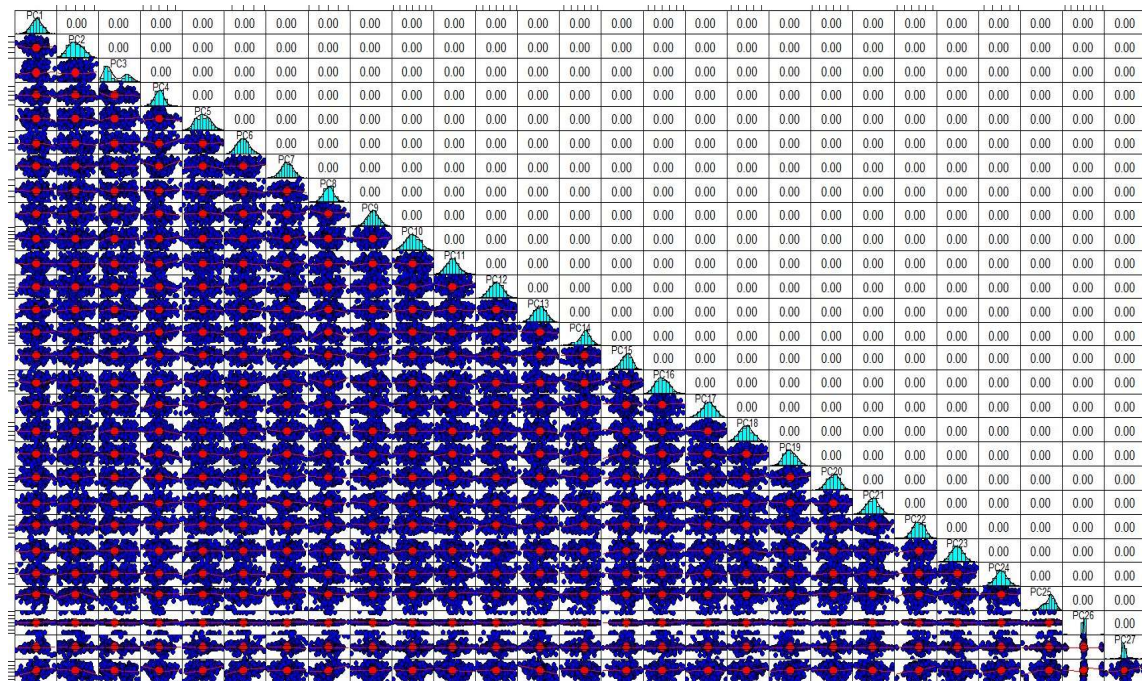


Figure 4: Orthogonal feature of all the variables after PCA

#### 4. MACHINE LEARNING CLASSIFIER

##### 4.1 Support Vector Machine (SVM)

This supervised ML technique is based on statistical learning and it tries to find a line that separates data into two groups [24]. This line is also known as a hyperplane that creates the biggest gap between two groups exploiting maximum possibilities [25]. A hyperplane is a robust feature of SVM that can be expressed as the set of data points ‘x’ satisfying

$$w \cdot x - b = 0 \quad (1)$$

parameter,  $b / \|w\|$  sets how far the hyperplane is positioned from the origin in the direction of the normal vector  $w$ .

##### Application of SVM on Raw Data

SVM is employed on the Target, a dependent variable by splitting the dataset with 80:20. Classification with a radial kernel having 520 support vectors, the raw dataset gives an accuracy

of 83.80% with 50.93% sensitivity and 99.34% specificity.

##### 4.2 Application of SVM on the Dataset with PCA

SVM applied on the same dataset using 80% data for training and 20% for testing. Again Target, as a dependent variable. Again, SVM employed using PCA on the Target, a dependent variable by splitting the dataset with 80:20. Classification with a radial kernel having 501 support vectors, dataset after PCA gives an accuracy of 90.19% with 72.22.93% sensitivity and 99.12% specificity.

##### 4.3 Performance Metrics

The receiver operating characteristic (ROC) and area under the curve (AUC) curve assess how well a classifier model performs at different threshold settings. It is also a graphical representation that shows how good the model is at telling things apart. A higher value of an AUC means that the model is better and intelligently recognizing ‘0s’ as ‘0s’ and ‘1s’ as ‘1s’, similar to how it is better at identifying patients with a disease from those with no disease. The ROC curve is a graphical representation that shows how

good a test is at finding things it's supposed to find in an ML model (TPR) and how many times it finds things it shouldn't (FPR).

Figure 5 and Figure 6 show the Receiver Operating Characteristics (ROC) results for the

components that were identified and extracted using the SVM classifier. The ROC curve is a tool used to analyze the performance of classifiers, curve (AUC). while the confusion matrix is a measure applied to assess the quality of the ROC curve by examining the area under the curve (AUC).

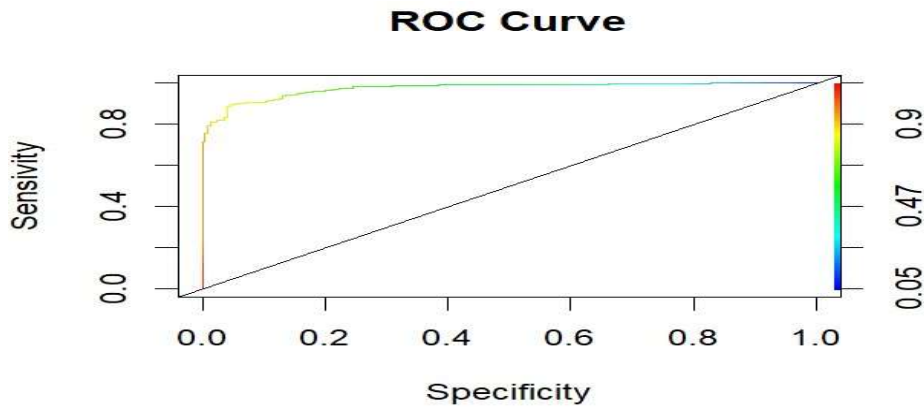


Figure 5: ROC curve of raw data PCA

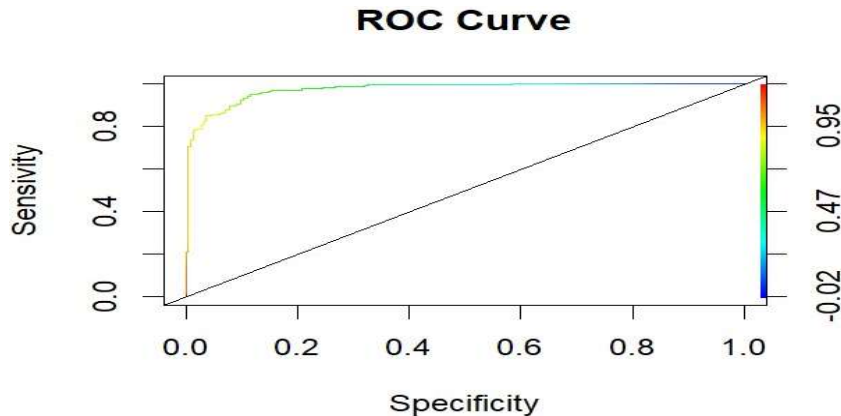


Figure 6: ROC curve after PCA

By observing the above figures, it may be analyzed in Table-1 Confusion Matrix that the model is doing better because the curves are above the line.

The AUC value is 97.43% found in raw data and 97.41% of normalized data after PCA. The model showed only a 0.02% difference in AUC value before PCA so the curve of Figure 5. is a little bit away from the boundary concerning Figure 6.

#### 4.5 Accuracy Analysis

A confusion matrix is an evaluating tool that is used to assess how well a classification system works by showing how often it gets things wrong or right. The accuracy of ML algorithms depends on four key components such as true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) from the confusion matrix (Table 2) of a machine learning model.



Table 1: Confusion matrix with different measurements of performance

		Predicted Values of raw dataset			
		673	0	1	
Actual	0	110 (TN)	3 (FP)		113
	1	106 (FN)	454 (TP)		560
		216	457		

Parameters Evaluation:

$$\text{Recall} = \frac{TP}{Actuals} = \frac{454}{560} = 81.10\%$$

$$\text{Accuracy} = \frac{(TP+TN)}{Total} = \frac{(110+454)}{673} = \frac{564}{673} = 83.80\%$$

$$\text{Precision} = \frac{TP}{Predicted\ yes} = \frac{454}{457} = 99.34\%$$

$$\text{Error rate} = 1 - \text{accuracy} = 1 - 83.80 = 16\%$$

Table 2: Confusion matrix with different measurements of performance

		Predicted Values after PCA			
		673	0	1	
Actual	0	154 (TN)	4 (FP)		158
	1	62 (FN)	453 (TP)		515
		216	457		

Parameters Evaluation:

$$\text{Recall} = \frac{TP}{Actuals} = \frac{453}{515} = 87.96\%$$

$$\text{Accuracy} = \frac{(TP+TN)}{Total} = \frac{(453+154)}{673} = \frac{607}{673} = 90.19\%$$

$$\text{Precision} = \frac{TP}{Predicted\ yes} = \frac{453}{457} = 99.12\%$$

$$\text{Error rate} = 1 - \text{accuracy} = 1 - 90.19 = 9.81\%$$

Where these features of a confusion matrix are denoted as:

TP (True Positives): The number of people identified correctly with heart disease.

TN (True Negatives): The number of people correctly identified as without heart disease.

FP (False Positives): The number of people incorrectly identified as having heart disease when they do not.

FN (False Negatives): The number of people incorrectly identified as not having heart disease.

5. RESULT

After using SVM, a machine learning approach with and without PCA on the dataset by splitting it in 80:20 ratio of training and testing we found that accuracy with PCA is better than without PCA. Accuracy was calculated by applying a confusion matrix of both algorithms as shown in Fig.6 and Fig.7 applied to calculate TP, TN, FP, and FN and it is concluded that SVM with PCA showed an accuracy of 90.49% shown in Table-3.

Table 3: Evaluation of various parameters in ML Approach with and without PCA

Evaluation parameter	ML Algorithm	
	SVM without PCA	SVM with PCA
Accuracy	83.80%	90.49%
Sensitivity	50.93%	72.22.93%
Specificity	99.34%	99.12%
AUC Score	97.43%	97.41%

This study leverages machine learning (ML) techniques, including Principal Component Analysis (PCA) and Support Vector Machine (SVM), to predict cardiovascular disease (CVD) using patient data taken from Rohilkhand Hospital for research purposes. Here an ML model, SVM with PCA produces better accuracy compared to SVM alone, achieving a higher accuracy rate of 90.49% v/s. 84.88%.

The research emphasizes the importance of early CVD detection and the role of machine



learning in healthcare analytics. Cardiovascular disease (CVD) is a worldwide health concern, PCA and SVM are important tools for analyzing and predicting medical data. SVM with PCA is superior in predicting CVD, Data preprocessing and performance metrics are essential.

## 6. CONCLUSION AND FUTURE SCOPE

To predict heart diseases, various algorithms are used to build machine learning models. The Evaluation parameters of a model are key factors in determining how well it can predict something correctly and identify heart problems. More specifically, accuracy depends on the insights that an algorithm learns from a set of data used for training and testing as well. We observed that SVM with PCA outperformed when the ML classifier was used for disease datasets with and without PCA.

The limitation of the study is that the dataset was taken from a specific region. This study can provide generalized outcomes if the data was collected from different regions of the country.

In the future, we can use these findings to predict and evaluate the correlation and the impact between the coronavirus vaccine and cardiovascular disease on a large dataset. Using different machine-learning techniques helps us to get even better predictions of heart diseases early. With the large data set, the accuracy of the various ML and DL models will be evaluated and it will provide a deeper insight into the performance of the deep learning model.

## 7. ACKNOWLEDGMENT

The authors would like to thank Rohilkhand Hospital, Shahjahanpur, Uttar Pradesh, India for providing data on patients to be used in this study. There was no monetary benefit to the hospitals for proving this data.

## REFERENCES

- [1] Mishra, S., Mallick, P. K., Tripathy, H. K., Bhoi, A. K., & González-Briones, A. (2020). Performance evaluation of a proposed machine learning model for chronic disease datasets using an integrated attribute evaluator and an improved decision tree classifier. *Applied Sciences*, 10(22), 8137.
- [2] Chicco, D., & Jurman, G. (2020). Machine learning can predict the survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC medical informatics and decision making*, 20(1), 1-16. WHO, "cardiovascular diseases," World Health Organization, 2021.
- [3] Olaniyi, A. S., Kayode, S. Y., Abiola, H. M., Tosin, S. I. T., & Babatunde, A. N. (2017). Student's performance analysis using decision tree algorithms. *Annals. Computer Science Series*, 15(1), 55-62.
- [4] 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), 1815518179.
- [5] Thummala, G. S. R., Baskar, R., & RimlonShibi, S. (2023, May). Prediction of Heart Disease using Random Forest in Comparison with Logistic Regression to Measure Accuracy. In *2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)* (pp. 1-5). IEEE.
- [6] Ziasabounchi, N., & Askerzade, I. N. (2014). A comparative study of heart disease prediction based on principal component analysis and clustering methods. *Turkish Journal of Mathematics and Computer Science (TJMCS)*, 16, 18.
- [7] 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.
- [8] Dun, B., Wang, E., & Majumder, S. J. C. S. (2016). Heart disease diagnosis on medical data using ensemble learning. *Comput. Sci*, 1(1), 1-5.
- [9] Singh, R. S., Saini, B. S., & Sunkaria, R. K. (2018). Detection of coronary artery disease by reduced features and extreme learning machine. *Clujul Medical*, 91(2), 166.
- [10] Yaghouby, F., Ayatollahi, A., & Soleimani, R. (2009). Classification of cardiac abnormalities using reduced features of heart rate variability signal. *World Applied Sciences Journal*, 6(11), 1547-1554.

- [11] Zhang, D., Zou, L., Zhou, X., & He, F. (2018). Integrating feature selection and feature extraction methods with deep learning to predict clinical outcome of breast cancer. *Ieee Access*, 6, 28936-28944.
- [12] Santhanam, T., & Ephzibah, E. P. (2013). Heart disease classification using PCA and feed forward neural networks. In *Mining Intelligence and Knowledge*
- [13] *Exploration: First International Conference, MIKE 2013, Tamil Nadu, India, December 18-20, 2013. Proceedings* (pp. 90-99). Springer International Publishing.
- [14] Dhankhar, A., & Jain, S. (2021). Prediction of disease using machine learning algorithms. *Smart and Sustainable Intelligent Systems*, 115-125.
- [15] Li, J. P., Haq, A. U., Din, S. U., Khan, J., Khan, A., & Saboor, A. (2020). Heart disease identification method using machine learning classification in ehealthcare. *IEEE access*, 8, 107562-107582.
- [16] Khan, S. N., Nawi, N. M., Shahzad, A., Ullah, A., Mushtaq, M. F., Mir, J., & Aamir, M. (2017). Comparative analysis for heart disease prediction. *JOIV: International Journal on Informatics Visualization*, 1(42), 227-231.
- [17] Bhatt, C. M., Patel, P., Ghetia, T., & Mazzeo, P. L. (2023). Effective heart disease prediction using machine learning techniques. *Algorithms*, 16(2), 88.
- [18] Hariharan, K., Vigneshwar, W. S., Sivaramakrishnan, N., & Subramaniaswamy, V. (2018). A comparative study on heart disease analysis using classification techniques. *International Journal of Pure and Applied Mathematics*, 119(12), 13357-13366.
- [19] Gárate-Escamila, A. K., El Hassani, A. H., & Andrés, E. (2020). Classification models for heart disease prediction using feature selection and PCA. *Informatics in Medicine Unlocked*, 19, 100330.
- [20] Weingessel, A., & Hornik, K. (2000). Local PCA algorithms. *IEEE Transactions on neural Networks*, 11(6), 1242-1250.
- [21] GÜNDOĞDU, S. (2022). Hepatitis C Disease Detection Based on PCA-SVM Model. *Hittite Journal of Science and Engineering*, 9(2), 111-116.
- [22] Slutsky, D. J. (2014). The effective use of graphs. *Journal of wrist surgery*, 3(02), 067-068.
- [23] Aladeyelu, A. C., & Adekunle, G. T. (2023). Predicting Heart Disease Using Machine Learning. *Machine learning*, 10(4).
- [24] Reddy, K. V. V., Elamvazuthi, I., Abd Aziz, A., Paramasivam, S., & Chua, H. N. (2021, July). Heart disease risk prediction using machine learning with principal component analysis. In *2020 8th International Conference on Intelligent and Advanced Systems (ICIAS)* (pp. 1-6). IEEE.
- [25] Narkhede, S. (2018). Understanding auc-roc curve. *Towards Data Science*, 26(1), 220-227.