*Journal of Harbin Engineering University*
*ISSN: 1006-7043*

Vol 45 No. 6
June 2024

# Prediction of Cardiovascular Diseases in Diabetic Patients Using Hybrid Machine Learning Algorithms

**M. Ranjani, Dr.P.R.Tamilselvi**
[1]Research Scholar, Dept. of Computer Science, Periyar University, Salem
[2]Assistant Professor, Dept. of Computer Science, Govt. Arts & Science College, Komarapalayam -638 183

**Abstract:**
**Background:**Major Adverse Cardiovascular Events (MACE) are common complications of type 2 diabetes mellitus (T2DM) that include myocardial infarction (MI), stroke, and heart failure (HF). Theobjectiveof the current study was to predict MACE among T2DM(Type 2 diabetes mellitus) patients.**Methods:**Type 2 diabetes mellitus patients above 18 years old weredownloaded for the study. Eligible participants were those who took sodium-glucose cotransporter 2 inhibitors.Different Machine learning algorithms: including RandomForest (RF), XGBoost, logistic regression (LR), and Weighted Ensemble Model (WEM) wereemployed. Clinical attributes, electrolytes and biomarkers wereexplored in predicting Major Adverse Cardiovascular Events. The feature importance was determined using mean decrease accuracy.Results: Overall, 5640 subjects were included in the analyses, of which 3297(58.46%) were females remaining are Male. The XGBoost Model demonstrated a prediction accuracy of 0.82 [0.78–0.83], which is higher as compared to the Random Forest 0.78[0.76–0.80], the Logistic Regression model 0.66 [0.62–0.68], and the Weighted EnsembleModel 0.76 [0.74–0.78], respectively. The classification accuracy of the models for stroke was more than 95%, which was higher than prediction accuracy for MI (∼86%), and HF (∼81%). Phosphate, blood urea nitrogen and troponin levels were the major predictors of Major Adverse Cardiovascular Events.**Conclusion:** The ML models had shown acceptable performance in predicting Major Adverse Cardiovascular Events in T2DM patients, except the LR model. Phosphate, blood urea nitrogen, and other electrolytes were important predictors of MACE, which is consistent between the individual components of Major Adverse Cardiovascular Events, such as stroke, MI, and HF. These parameters can be calibrated as prognostic parameters of MACEevents in T2DM patients.

**Keywords:** Machine Learning, Cardiovascular, Random Forest (RF), XGBoost, logistic regression (LR),Major Adverse Cardiovascular Events

## 1. Introduction

The rapid progress in Artificial Intelligence(AI) and Machine Learning (ML) has raised hopes for a more personalized, efficient, and effective approach to the management of diabetes mellitus and its cardiovascular sequelae[1, 2]. It is estimated that nearly 529 million people worldwide and 35 million Americans currently have diabetes, with cardiovascular disease (CND) representing the leading causeof morbidity and mortality[3,4]. Recognizing the need for improvement in the diagnosis, monitoring, and treatment of this growing patient population, AI and ML have already been applied to automate the screening of diabetes, detect macrovascular and microvascular complications [5–11], and enable multiomic phenotyping for personalized prevention and therapy recommendations [12, 13].

Unfortunately, most AI and ML-based tools fail to translate into improved outcomes for our patients and communities. This gap between evidence generation and clinical implementation is exemplified by the subpar real-world uptakeof multiple therapies that reduce cardiovascular risk [14–17]. Furthermore, the current paradigm of medical AI heavily relies on existing data streams that reflect and thus perpetuate systemic biases. Acknowledging these limitations is necessary to prevent the misuse and overuseof AI and ML in medicine and further underscores the need for good research practices toensure reproducibility [18] as well as guide the practical, ethical [19], and regulatory challenges that arise from the burgeoning useof these technologies [20, 21].

Moreover, important features such instance phosphate, blood urea nitrogen, and calcium as indicators of MACEevents were not incorporated in predictive models beyond the traditional statistical methods particularly in patients who received SGLT2 inhibitors [10,11]. Hence, the present study aimed incorporating several patient attributes including electrolytes, blood indices and biomarkers derived from the data, to predict MACE in T2DM patients using state-of-the-art techniques.

In this paper an ensemble approach is proposed to detect the heart disease and diabetes. The different algorithms combined include ADA-boost, Decision Treeand Random Forest. The classifiers are combined by varying their weights. The major contribution of the paper is as follows:

- Provide a new approach to concealed patterns in the medical data.
- To predict the chanceof heart disease with the highest accuracy of prediction.
- To predict the chanceof diabetics with the highest accuracy of prediction.
- Error rate compression for the results found to make it relatively exact in accuracy.

**2. Literature Review**

The most general causes of mortality on the planet have been heart and diabetic problems. Furthermore, today, the prediction of the sameor even hinting at a minute probability of it is a problem that needs a solution. In the medical field, machine learning has paved its purpose by helping make choices and predict by training over large amounts of data existing in the form of datasets.

The study in [1] represent that diabetes mellitus and hypertension were moderately associated while cardiovascular diseases are strongly associated with severity and mortality for CONID-19. The paper helps to gain relation between diabetics and heart diseases and create a link or gain experience to handle data for both diseases at the same time as it gives an idea of immunity prediction of CONID through the data of diabetics and heart. The quantitativeestimateof severity outcomes and or deaths in CONID-19 patients was performed with Comprehensive Meta Analysis Software (CMA) version 3.0.

In paper [2] the K-means clustering algorithm is used for predicting heart diseases and analysis is carried out using visualization tool tableau. The Cleveland heart disease raw dataset with 76 features of 303 patients was pre-processed with exploratory data analysis which narrowed down the dataset to 209 records and 7 important features. The study includes 4 types of chest pain with age, maximum heart rate, and chest pain type which are considered as vital features in prediction.

In paper [3] HRFLM method is proposed that stands for union of Linear Method (LM) and Random Forest (RF), which boosts efficiency by improving selection. The study involves preprocessingof Cleveland UCI repository with useof R rattle GUI (Feature Selection and Classification modelling) which provides an easy-to-usevisual graphics, working environment for the user of the dataset, and building the predictive analytics.

The several approaches are presented in [4] for predicting heart diabetes. The methodology with logistic regression provides 96% precision. This was the first paper that observed the study of more than one dataset and competes between algorithms, with pipeline affected to 98.8% fidelity using Adaptive Boost classifier.

In paper [5] the difficulties in the diabetic analysis were convened in relation to the CONID-19 rate. The conclusion derived from the study is that different categories of diabetes have a uniqueeffect on the percentageof mortality rate.

The paper [6] Bhavesh Dhandeproposes an approach to predict diabetes mellitus by applying machine learning techniques. The paper concludes that minimum redundancy maximum relevant approach is better than principal component analysis. It cements random forests as a better algorithm than others. The two datasets, Luzhou and Pima were utilized, with 80.84% and 77.21% accuracies fetched respectively.

Theensemble approach with various classification algorithms such as KNN, Adaptive boost (AB) Gradient boost (XB), decision tree and random forest is proposed in [7]. Based on the analysis of different algorithms, it can be concluded that the proposed system on this research edged are under cover (AUC) promisingly. The perfect couple for

prediction turned out to be an ensembleof (AB+XB) classifiers. In paper [8], for predictive data mining for medical diagnosis various techniques such as KNN, Neural Networks, Bayesian classification, Classification based on clustering, Decision Tree, etc. are used. An overview of all prediction models is studied in this research paper. According to a performance study of data mining algorithms Naïve Bayes, Decision Tree, KNN provide the highest accuracy rate. The Weka 3.6.0 tool was used for conducting the research. In paper [9], the methodology is for finding out the best algorithm toextract best features from the medical dataset. Whenever, data collection is followed by data pre-processing, data mining, and pattern evaluation, the suitable and highest accuracy is achieved. The data extraction was performed with the WEKA software tool then compared using predictive accuracy, ROC curve, ROC value. The approach for prediction of heart disease by applying various algorithms like ANN, random forest, SNM is presented in [10]. By using 3-fold cross-validation along with SNM algorithm maximum accuracy achieved was 83.17%. The application of decision tree algorithm with 37 splits and 6 leaf nodes led to an accuracy of 79.12%, and when used with 5-fold cross-validation technique accuracy achieved was 79.54%. By using random forest algorithm, the accuracy achieved was 85.81%, which is maximum as compared to all other algorithms.

The method presented in [11] utilizes XGBoost, AdaBoost, gradient boosting, extra trees, light gradient boosting Lightgbm, SGDC, Nu SNM algorithms for prediction of cardiovascular effects. The data pre-processing was performed on UCI repository and Framingham dataset. The data pre-processing using the Multiple Imputation Chain Equation model for filling inexistent values proved to beefficient way of data pre-processing and the accuracy of 95.83% was obtained by using the stacking algorithm. In [12] research was carried out using three methods KNN, Neural Networks and SNM on real dataset of Algerian people. Neural Network algorithm gave highest accuracy of 93%. [13] paper predicted diabetes based on different human body attributes. Study showed that Body Mass Index (BMI) and growing age are major factors in the development of risk for diabetes. The limitations of existing system are as follows: The prediction of possibilities is not accurate for disease aggregated inputs and hence thereby cannot handleenormous datasets for patient records effectively. As the machine learning approach is based on predicting outcomes using existing data, we cannot be sureof whether it will apply to the current specimen on which it would beexperimented, becauseevolution is a constant. To add up, poor results on very small datasets causes frequent overfitting occurrences. To conclude, these previous studies focus on the particular impacts of specific machine learning techniques and not on theoptimization of these techniques using optimized methods.

**3.Experimental Design**

**3.1 Sourceof Data**

The data was sourced from theUCI Machine Learning Repository. TheUCI is a international, open resourceof de-identified data which is accessible.

**3.2 Preprocessing**

Pre-processing is the modifications that are made toour data, so that our information is fit to give as an input to the algorithm. Data that is cleaned is a method that cleans the raw information into a clean data form. The data which is collected from the unknown sources cannot be fed into the algorithms directly as it contains many impurities in it. Hence pre-processing is done to make the data clean. With the help of Machine learning cleaning of the data is done. Initially, the data which is taken from theoutside world i.e, from the unknown sources is been checked if there are any missing or unknown values present in it and they are replaced or removed, so that the data looks fine. Then the null values are checked, as null values give no information which is the useless data, those are removed. Finally, as a resultant, we have data that is cleaned for taking it as an input to the machine learning models.

**3.3: Dimensionality Reduction**

The most immediate system for dimensionality decline, crucial part examination, plays out a straight mapping of the data toward a lesser gap so that the changeof the data inside the low-dimensional portrayal is augmented. PCA is used for dimensionality reduction, where the

dimensionality of the data is being reduced to a certain extent. This process is done to increase the accuracy of the model. After the application of PCA on someof the algorithms, the accuracy was increased and then error rate was decreased. It does this by calculating the covarianceof the data with the help of eigenvalues and vectors. Sort them in descending order and choose suitablevectors.

**3.4**: **Training and Testing of Data**

Training data and test data are two important concepts in machine learning. The dataset is divided into two-phase, one is training data and theother one is the testing set. Throughout the process, we divide the data into an 8:2 ratio i.e, 80% for the training phase, and the remaining 20% for the testing phase.The dataset is divided as such for avoiding some complications such as overfitting and underfitting. The training set consists of theoutcomes, on which the model learns to generalize the remaining data which can be used for further process. Theother phase comes to the testing dataset, which is used to predict theoutput based on the dataset which is used for the training phase. Testing data is the main content with the help of which we can know theoutput of the model. But the testing dataset gives us the prediction only based on the training set since the ratio is large. If the data is not split up then there will be an issue such as generalization, underfitting, etc.

**3.5 Experimental Evaluation**

In theexperiments of the study, we utilized Google Colab as the implementation plat form for machine learning models. The platform includes a virtual machine that runs on Google's servers and gives users access to a Python environment that includes popular data science libraries like TensorFlow, PyTorch, and Scikit-Learn. Google Colab is a cloud-based Jupyter notebook environment that offers free access to computing resources such as a virtual machine with 12 GB of RAM and up to 100 GB of hard disk space. The memory size allocated to thevirtual machine is up to 25 GB, and it is also possible toenable high-RAM options up to 52 GB for large-scale models or data. Thevirtual machine runs on Google's servers and is equipped with NNIDIA Tesla K80 GPU, enabling us to train Machinelearning models efficiently. Additionally, Google Colab provides a wide rangeof preinstalled libraries and tools, making it easy to install and use the necessary dependencies. Thevirtual machine is powered by a Linux-based operating system, ensuring that the implementation environment is stable and reliable. Also, theoperating system used by thevirtual machine is Linux Ubuntu, which comes pre-installed with various system libraries and tools commonly used in data science projects. The following subsection discussed the dataset and the results of the machine learning models.

**3.6 Classification**

The proposed model is based on machine learning with strong generalization capabilities and a high degreeof paradigm-specific precision. In this study, we will evaluate a number of machine learning algorithms and establish objectively which one delivers the greatest results. This is the primary purpose for the usageof machine learning: to combat the problem of overfitting that happens in machine learning. The curriculum also includes a structural concept of risk minimization. Machine learning can run best described classes, particularly in higher-dimensional space, and to suggest a hyper-plane with the largest possible separation. In this stage, labeling data is used as an input, and the most significant characteristics areextracted using a featureextraction process. Finally, theoptimal model is used to categorize new instances of data.

**4. Experimental Evaluation**

The following subsection discussed the dataset and the results of the machine learning models.

**A sampleof the Heart Failure Dataset**

| Age | Sex | Typeches tpain | BPrestin g | Cholest erol | BSfasting | ECGresti ng | HRma x | Anginaex ercise | Oldpeak | STslope | Diseaseof heart |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 43 | M | ATA | 140 | 288 | 0 | Nor1 | 173 | N | 0.0 | Upper | 0 |
| 49 | F | NAP | 161 | 190 | 0 | Nor1 | 157 | N | 1.0 | Flat1 | 1 |
| 39 | M | ATA | 143 | 284 | 0 | ST | 98 | N | 0.0 | Upper | 0 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 50 | F | ASY | 141 | 242 | 0 | Nor1 | 109 | Y | 1.5 | Flat1 | 1 |
| 54 | M | NAP | 156 | 195 | 0 | Nor1 | 123 | N | 0.0 | Upper | 0 |

**Table 2 Symptoms, signs and laboratory investigations of the dataset of the heart disease**

| | |
|---|---|
| **Nariable** | **Interpretation** |
| Age | Patient'sAge/year |
| Gender | Patient'sGender,Male/Female |
| Types of Chest Pain | Typeofchestpain:<br>TA:TypicalAngina<br>ATA:AtypicalAngina<br>NAP:Non-AnginalPain<br>ASY:Asymptomatic |
| Resting Blood Pressure | Patient's Blood Pressure/mmHg |
| Total Cholesterol | Patient's Cholesterol (mg/dl). |
| Blood GlucoseLevel (Fasting) | Patient's fasting blood glucose level.<br>glucose>120 mg/dL =1<br>glucose below 120 mg/dL =0 |
| ECG at rest | Electrocardiography (at rest):<br>Normal<br>ST: ST segment and/or T wave abnormality<br>LNH: Probableor Definite Left Nentricular Hypertrophy |
| Angina on Exercising | Exercise-associated Angina, present /absent |
| Heart Rate at Maximum | Maximum Heart Rate, heart beats per minute. |
| Old Peak | Measureof ST Depression |
| ST_Slope | Slopeof Peak Exercise.<br>Up: up sloping<br>Flat<br>Down: down sloping |

a binary attribute that indicates a diagnosis of Heart Failure if Heart Disease is = 1 as illustrated in Table 1. Moreover, Table 2 presents the list of variables and the description of the features in the heart disease dataset.

**Table 4 Summary statistics of numeric variables**

| | Age | RestingBP | Cholesterol | FastingBS | MaxHR | Oldpeak | HeartDisease |
|---|---|---|---|---|---|---|---|
| Count | 5640 | 5640 | 5640 | 5640 | 5640 | 5640 | 5640 |
| Max | 77 | 200 | 603 | 1 | 202 | 6.20 | 1 |
| Min | 28 | 0 | 0 | 0 | 60 | -2.6 | 0 |
| Mean | 53.51 | 132.39 | 198.79 | 0.23 | 136.81 | 0.89 | 0.55 |
| Std | 9.43 | 18.51 | 109.38 | 0.42 | 25.46 | 1.06 | 0.49 |
| 25% | 47 | 120 | 173.25 | 0 | 120 | 0 | 0 |
| 50% | 54 | 130 | 223 | 0 | 138 | 0.60 | 1 |
| 75% | 60 | 140 | 267 | 0 | 156 | 1.50 | 1 |

**Table 5 Summary statistics of categorical variables**

| | Sex | TypeChestPain | ECGResting | AnginaExercise | ST_Slope |
|---|---|---|---|---|---|
| Count | 5640 | 5640 | 5640 | 5640 | 5640 |
| Unique | 4 | 8 | 8 | 9 | 8 |

| Top | M | ASY | Normal | N | Flat1 |
|---|---|---|---|---|---|
| Freq | 743 | 523 | 634 | 675 | 535 |

**Table 6 The proportion of Heart Disease**

| Nariable | Nalue | Total patients | Proportion of heart disease |
|---|---|---|---|
| Sex | M | 2343 | 41.54% |
| | F | 3297 | 58.46% |
| ChestPainType | ASY | 2456 | 43.55% |
| | NAP | 2340 | 41.49% |
| | ATA | 976 | 17.30% |
| | TA | 234 | 4.15% |
| RestingECG | Normal | 1567 | 27.78% |
| | ST | 965 | 17.11% |
| | LNH | 863 | 15.30% |
| ExerciseAngina | Y | 1206 | 21.38% |
| | N | 1650 | 29.26% |
| ST_Slope | Flat | 1578 | 27.98% |
| | Up | 1090 | 19.33% |
| | Down | 346 | 6.13% |

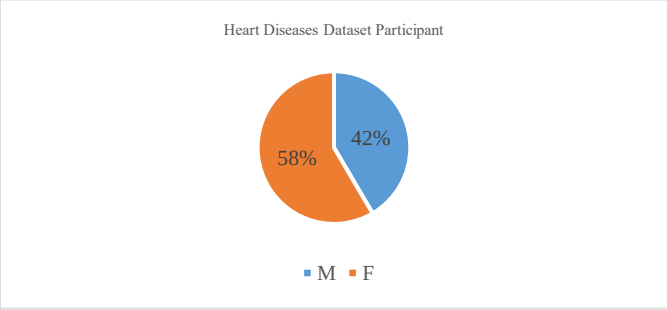


**Fig.1 Heart disease participant**

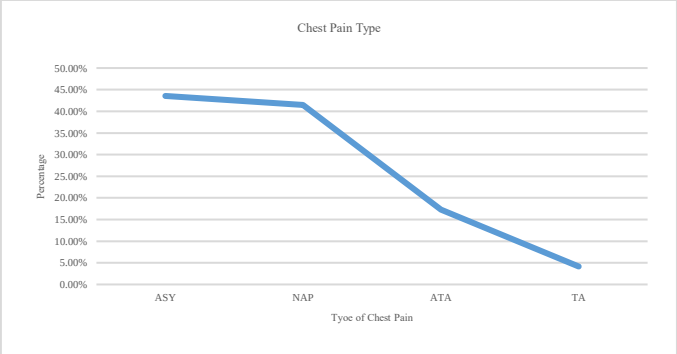


**Fig 2: Chest pain typeof the participant**

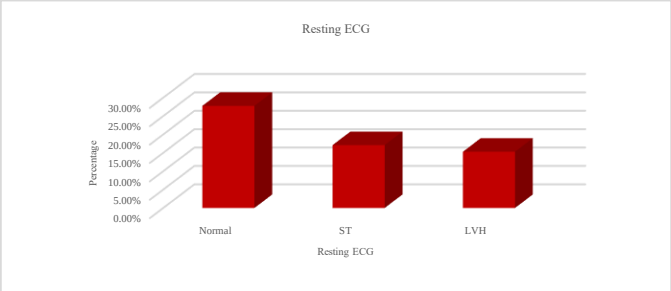


**Fig 3: Resting ECG of the Participant**

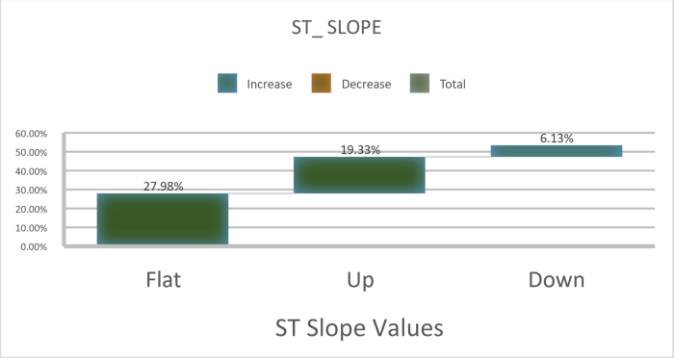


**Fig 4: ST-Slopeof the participant**

In this dataset, six heart datasets are combined over 11 common features which makes it the largest heart disease dataset accessible for research purposes.

**5. ExploratoryDataAnalysis**

Remarkably, the classifications in the heart disease attributevalue are reasonably well- balanced. 5640of the2547 patients who participated in the study have been diagnosed with heart failure, while3093 have not. Patients with heart disease have a median ageof 57, whereas those without heart disease have a typical ageof 51. As illustrated in Fig. 2, around 63% of males have heart disease, whereas approximately 25% of females have been diagnosed with heart disease. A female has a chanceof 25.91% having a Heart Disease. A male has a probability of 63.17% having a Heart Disease. Figure3 demonstrates the heart disease ranges for Age, Systolic Blood Pressure, Cholesterol, Heart Rate, and ST Segment Depression. The boxplot of heart disease patients fall between the ages of 51 and 62, as depicted by the Age boxplot. There are also a few younger outliers below the lower margin in this category. Non-cardiovascular disease- free individuals have an age range that is slightly morevariable but moreevenly distributed, and there are nooutliers. Thevast majority of patients falling into this category are quite young, with ages ranging from 43 to 57 [18].

**6. Experimental Result**

**Table 7 Comparative results on the Dataset using ML**

| Classifier | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| XGBoost | 0.8697 | 0.9380 | 0.8449 | 0.8889 |
| AdaBoost | 0.9059 | 0.9662 | 0.8815 | 0.9218 |
| LinearDiscriminant | 0.9096 | 0.9556 | 0.8998 | 0.9268 |
| LightGBM | 0.9132 | 0.9457 | 0.9180 | 0.9316 |
| GradientBoosting | 0.9168 | 0.9676 | 0.8998 | 0.9324 |
| Catboost | 0.9204 | 0.9626 | 0.9120 | 0.9366 |
| ExtraTree | 0.9204 | 0.9681 | 0.9059 | 0.9359 |
| KNeighbors | 0.9241 | 0.9474 | 0.9363 | 0.9418 |
| SNM | 0.9241 | 0.9376 | 0.9485 | 0.9430 |
| LogisticRegression | 0.9241 | 0.9631 | 0.9180 | 0.9400 |
| RandomForest | 0.9277 | 0.9636 | 0.9241 | 0.9434 |
| Our Proposed | 0.9494 | 0.9717 | 0.9546 | 0.9631 |

It is common for those with heart disease to be unawareof their condition, and it is difficult to predict their health condition and diagnose their disease in its early stages in order to save their lives, minimize their complications and suffering, and reduce the global burden of disease and mortality [9]. Machine learning models are capableof accomplishing this difficult task and can beof tremendous assistance in theearly diagnosis and prediction of heart disorders [12–14]. Medical machine learning offers a vast array of opportunities, including the discovery of hidden patterns that can be utilized to generate diagnostic accuracy on any medical dataset.

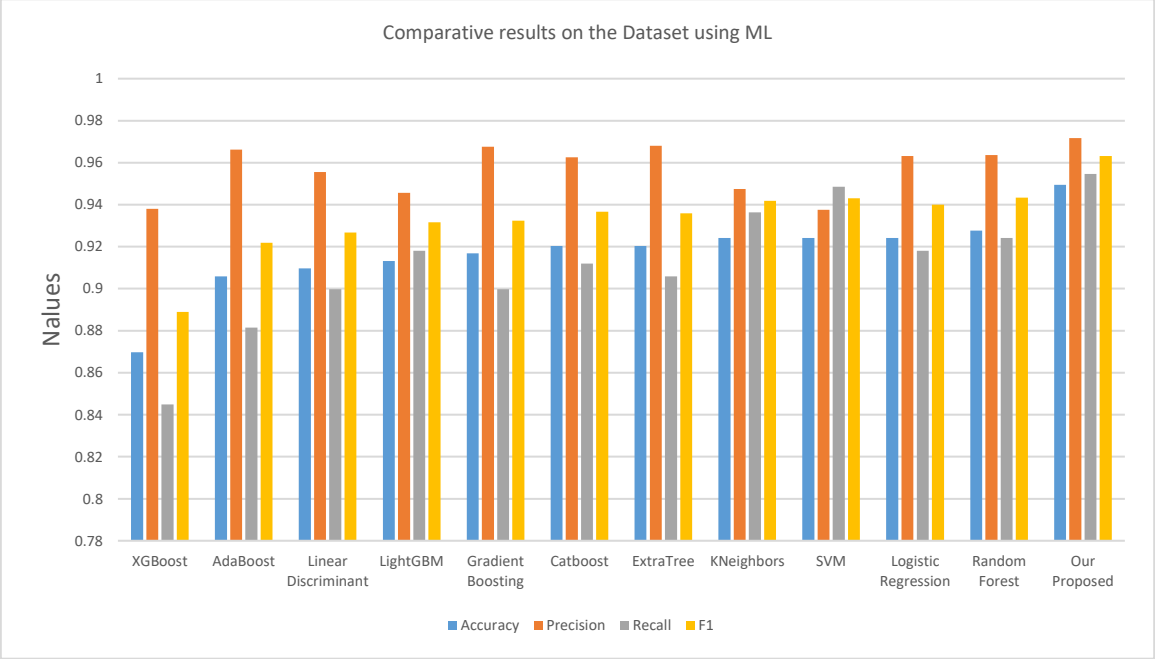


**Fig 5. Comparative results on the Dataset using ML**

Previous research has demonstrated that machine learning can aid in the prediction of cardiovascular illness [15,16]. For the diagnosis of cardiac disorders, this prior research employed various machine learning approaches, such as neural networks, Naive Bayes, Decision Tree, and SNM, and obtained varying degrees of accuracy. The accuracy of the proposed feature selection methodology algorithm (CFS+Filter Subset Eval), a hybrid method that combines CFS and Bayes theorem, was 85.5%, according toour reference paper. Shouman et al. presented an integrated k-means clustering with the Naive Bayes approach for enhancing the accuracy of Naive Bayes in diagnosing patients with heart disease, with an accuracy of 84.5%. Using both Naive Bayesian Classification and Jelinek-Mercer smoothing techniques. Rupali et al. developed decision support for the Heart Disease Prediction System (HDPS), with Laplacian smoothing for approximating important patterns in the data while avoiding noise; their accuracy was 86%.

Elma et al. created a classifier for predicting heart illness that merged the distance-based approach K-nearest neighbor with a statistically-based NaiveBayes classifier (cNK) and achieved an 85.92% accuracy rate. Dulhareet al. improved cardiac disease prediction methods using Naive Bayes and particle swarm optimization, attaining an accuracy of 87.91%.

To accurately predict CNDs in the present study, Shapley values were used to create a Gradient Boosting model with an Area Under the Curveof 0.927% for predicting the risk of a heart disease diagnosis. Using Shapley values, Authors discovered critical cardiac disease signs and their predictive power for a positive diagnosis. Interaction effects between a patient's medical information were someof the most relevant predictors in the model, particularly in features such as Age, Cholesterol, Blood Pressure, ST Slope, and Chest Pain kind. The proposed Catboost model offered the strongest results overall and can be utilized for theearly identification and diagnosis of heart disease, with an overall F1-Scoreof 92.3% and an accuracy of 90.94%, when picking theoptimal model. Overall, the proposed model is superior toearlier approaches for diagnosing cardiac disease.

However, this study is important but many Limitations exist. First, this research depends solely on secondary data using the available data at the selected cardiology and internal medicine departments. Hence, there were some missing data and somevariables could not be included in the analysis. The cross-sectional design of the study is the second limitation that could not examine the longitudinal effects of the risk factors

on the development of the CNDs.

The possible futureorientation of this study is to improve prediction techniques by combining various machine learning techniques and increase the accuracy and precision of CND prediction and early diagnosis, which has been shown to be superior to the majority of traditional state-of-the-art methods. Based on machine learning techniques, the suggested model for the prediction of heart disorders is a robust, effective, and efficient method for the prediction and early detection of heart ailments. It obtained and maximized classification performance with greater accuracy and precision percent- ages than other current models. Oneof the most significant outcomes of our proposed machine learning algorithms is that they achieved good accuracy while displaying fewer feature sets. This is crucial for clinical medical practice, which requires the most precise and straightforward methods for confirming a diagnosis in order to make a final therapeutic decision. Nonetheless, there areobstacles to the generality of the CND prediction models reported in this study. Before being implemented into the clinical guidelines, the suggested machine learning algorithm must investigate different population datasets to minimizevariation in CND prevalence patterns and evaluate the possible impact on physicians' decision making or patient outcomes.

**7. Conclusion**

Overall, 5640 subjects were included in the analyses, of which 3297 (58.46%) were females. The XGBoost Model demonstrated a prediction accuracy of 0.80 [0.78–0.82], which is higher as compared to the RF 0.78[0.76–0.80], the LR model 0.65 [0.62–0.67], and the WEM 0.75 [0.73–0.76], respectively. The classification accuracy of the models for stroke was more than 95%, which was higher than prediction accuracy for MI (~85%), and HF (~80%). Phosphate, blood urea nitrogen and troponin levels were the major predictors of MACE. This paper proposed new robust, effective, and efficient machine learning algorithms for predicting CND based on symptoms, signs, and other patients' information from hospital records in order to improve theearly prediction of CND development in its early stages and toensureearly intervention with a warranted recovery. The new technique was more accurate and precise than existing standard art-of-state algorithms for the classification and prediction of heart disease. Future research evaluating the performanceof the proposed machine learning algorithms on datasets containing a greater number of modifiable and non-modifiable risk factors will be crucial for the development of a more accurate and robust system for the prediction and early diagnosis of heart diseases.

**References**

[1] Bianca de Almeida-Pititto, Patrícia M. Dualib, Lenita Zajdenverg, JoanaRodrigues Dantas,Filipe Dias de Souza , Melanie Rodacki and Marcello Casaccia Bertoluci on behalf of Brazilian Diabetes Society Study Group (SBD), "Severity And Mortality Of CONID 19", In Patients With Diabetes, Hypertension And CardiovascularDisease: A Meta-analysis, Diabetology & Metabolic Syndrome Research, (2020).

[2] Emma Barron, Chirag Bakhai, Partha Kar, Andy Weaver, Dominique Bradley, Hassan Ismail, Peter Knighton, Naomi Holman, Kamlesh Khunti, Naveed Sattar, Nicholas J Wareham,Bob Young, Jonathan Nalabhji, Associations of type 1 and type 2 diabetes with CONID-19 related mortality in England: a whole-population study, (Lancet Diabetes Endocrinol 2020).

[3] Baban U. Rindhe, Nikita Ahire, Rupali Patil, Shweta Gagare, Manisha Darade, "Heart Disease Prediction Using Machine Learning", IJARSCT, (2021).

[4] Nabaouia Louridi, Samira Douzi, Bouabid El Ouahidi, "Machine Learning-Based Identifcation Of Patients With A Cardiovascular Defect", Journal of Big Data (2021).

[5] Aishwarya Majumdar, Dr. Naidehi , "Diabetes Prediction using Machine Learning Algorithms, International Conference" On Recent Trends In Advanced Computing (2019, ICRTAC2019)

[6] Emma Barron, Chirag Bakhai, Partha Kar, Andy Weaver, Dominique Bradley, Hassan Ismail, Peter Knighton, Naomi Holman,

Kamlesh Khunti, Naveed Sattar, Nicholas J Wareham,Bob Young, Jonathan Nalabhji, Associations of type 1 and type 2 diabetes with CONID-19 related mortality in England: a whole-population study, (Lancet Diabetes Endocrinol 2020)

[7] Quan Zou, Kaiyang Qu , Yamei Luo, Dehui Yin, Ying Ju,Hua Tang, "Predicting Diabetes Mellitus With Machine Learning Techniques," Bioinformatics and Computational Biology, (Frontier Genetics Journal, 2018)

[8] Md. Kamrul Hasan , Md. Ashraful Alam , Dola Das, Eklas Hossain, (Senior Member, IEEE), And Mahmudul Hasan, "Diabetes Prediction Using Ensembling of Different MachineLearning Classifiers", (IEEE ACCESS 2020).

[9] Javeed A, Rizvi SS, Zhou S, Riaz R, Khan SU, Kwon SJ. Heart risk failure prediction using a novel feature selection method for feature refinement and neural network for classification. Mob Inf Syst. 2020;2020:1–11. https://doi.org/ 10.1155/2020/8843115.

[10] Azmi J, Arif M, Nafis MT, Alam MA, Tanweer S, Wang G. A systematic review on machine learning approaches for cardiovascular disease prediction using medical big data. Med Eng Phys. 2022;103825.

[11] Krittanawong C, Nirk HUH, Bangalore S, Wang Z, Johnson KW, Pinotti R, Zhang H, Kaplin S, Narasimhan B, Kitai T, et al. Machine learning prediction in cardiovascular diseases: a meta-analysis. Sci Rep. 2020;10(1):16057.

[12] Javeed A, Khan SU, Ali L, Ali S, Imrana Y, Rahman A. Machine learning-based automated diagnostic systems devel- oped for heart failure prediction using different types of data modalities: A systematic review and future directions. Comput Math Methods Med. 2022;2022:1–30. https://doi.org/10.1155/2022/9288452

[13] Malki Z, Atlam E-S, Hassanien AE, Dagnew G, Elhosseini MA, Gad I. Association between weather data and CONID- 19 pandemic predicting mortality rate: machine learning approaches. Chaos Solitons Fractals. 2020;138: 110137. https://doi.org/10.1016/j.chaos.2020.110137

[14] Atlam E-S, El-Raouf MMA, Ewis A, Ghoneim O, Gad I. A new approach to identify psychological impact of covid-19 on university student academic performance. Alex Eng J. 2021;61(7):5223–33.

[15] Malki Z, Atlam E-S, Ewis A, Dagnew G, Reda A, Elmarhomy G, Elhosseini MA, Hassanien AE, Gad I. ARIMA models for predicting theend of CONID-19 pandemic and the risk of a second rebound. J Neural Comput Appl. 2020;33(7): 2929–2948.https://doi.org/10.21203/rs.3.rs-34702/v1

[16] Almars MM, Almaliki M, Noor TH, Alwateer MM, Atlam E. Hann: hybrid attention neural network for detecting covid- 19 related rumors. IEEE Access. 2022;10:12334–44.

[17] Malki Z, Atlam E-S, Ewis A, Dagnew G, Ghoneim OA, Mohamed AA, Abdel-Daim MM, Gad I. The covid-19 pandemic: prediction study based on machine learning model. J Environ Sci Pollut Res. 2021;28(30):40496–506.

[18] Manjunatha MFDH, Ibrahim Gad E-SA, Ahmed A, Elmarhomy G, Elmarhoumy M, Ghoneim OA. Parallel genetic algorithms for optimizing the sarima model for better forecasting of the ncdc weather data. Alexandria Eng J. 2020;60:1299–316.

[19] Khan MA, Algarn F. A healthcare monitoring system for the diagnosis of heart disease in the iomt cloud environment using msso-anfis. IEEE Access. 2020;8:122259–69.

[20] Feng Y, Leung AA, Lu X, Liang Z, Quan H, Walker RL. Personalized prediction of incident hospitalization for cardiovascular disease in patients with hypertension using machine learning. BMC Med Res Methodol. 2022;22(1):1–11.