VFAST Transactions on Software Engineering
http://vfast.org/journals/index.php/VTSE@ 2023, ISSN(e): 2309-3978, ISSN(p): 2411-6246
Volume 11, Number 2, April-June 2023 *pp: 140-148*

Optimized Classification of Cardiovascular Disease Using Machine Learning Paradigms

Fouzia Kanwal[1], Mr. Kamran Abid[1], Muhammad Sajid Maqbool[2], Naeem Aslam[1], Muhammad Fuzail[1]

[1]Department of computer science, NFC Institute of Engineering and Technology Multan, Pakistan
[2]Department of computer science, Bahauddin Zakariya University Multan, Pakistan
*Corresponding author email: fouzia53391@gmail.com

ABSTRACT

Nearly 19 million people die each year from cardiovascular and chronic respiratory diseases, which are a global threat. It is necessary to address the causes of these diseases because of the high death rate. The investigation uncovered a number of causes, but the inability to forecast these diseases symptoms is by far the most significant. In this work, we developed a method for anticipating these diseases crucial symptoms, which will aid in early disease diagnosis and allow patients to begin treatment. This research will introduce a new computational medicine research using machine learning (ML) paradigms to forecast cardiovascular disease (CVD). Data were processed by methods in sequence with various parameters. different models created that predicts CVD risk based on individual age, gender, ethnicity, body mass etc., and lifestyle factors. The research will also focus on performing complete comparison of ML models. We will apply Five ML based algorithems such as Decision Tree (DT), K-Nearest Neighbors (KNN), Naïve Bayes (NB), XGBOOST and Random Forest and evaluate these models on the basis of Training and Testing and also calculated the Presicion Recall and F1-Score for each model. Naïve Bayes and XGBOOST Classifier perform better with accuracy of 92.31 and 92.34 percent as compared to other models.

Keywords:
Cardio Vascular, Cardio Vascular Disease, Data Science, Student Performance Prediction, Machine Learning, Deep Learning

JOURNAL INFO
HISTORY: Received: May 24, 2023 Accepted: June 26, 2023 Published: June 30, 2023

1 INTRODUCTION

The World Health Organization (WHO) found that heart attacks and strokes are responsible for 17.5 million deaths worldwide. Furthermore, low- and middle-income nations account for the majority of the deaths from cardiovascular diseases, which account for almost 75% of all deaths worldwide. The third most prevalent illness worldwide and a major cause of death is heart disease, sometimes referred to as cardiovascular disease. CVDs encompass conditions that affect the heart, such as coronary heart disease, cerebrovascular illness, rheumatic heart disease, and others. bad eating habits, inactivity, smoking, and drinking too much alcohol often cause them. In this blog post, we will be using machine learning and deep learning paradigms to forecast cardiovascular disease rates on a global scale [1]. “Heart failure is a condition where the heart loses its ability to pump enough oxygen-rich blood to meet the body's needs”. This leads to a retention of fluid, breathing problems, and swelling in the lower legs (peripheral edema). Central edema happens when the brain gets too much fluid and can result in mental confusion or drowsiness. Heart failure affects men and women equally and can happen at any age. Symptoms vary from person to person; however, they can include shortness of breath, chest pain, unusual weight gain or loss, fatigue, weakness or numbness in one arm or leg; dizziness or fainting episodes; and needing to urinate more often than normal. Heart diseases are the most significant issue faced by people. It is caused by high cholesterol, high blood pressure, and other chronic diseases. Therefore, a large number of Americans have reduced their risk factors for heart disease by eating healthier foods and maintaining physical activity. As most people do not understand the severity of heart disease, they often miss or forget to check their hearts. Similarly, people who feel symptoms of cardiac issues or discomfort do not treat them which can lead to an early stage of heart failure associated with chronic symptoms like tiredness and fatigue. In this work we look at why understanding the signs and symptoms of heart disease is important to catch any disease earlier on. Treating cardiac issues early means more chances of survival, less medication, and no downtime because it has been treated while it’s still early in its course [2].

In this Section, we introduce the topic of our research and provide background information on the importance of studying cardiovascular disease (CVD) using machine learning techniques. We outline the research questions that we aim to address and the objectives of our study. We also provide an overview of the structure of the thesis, including the organization of the subsequent chapters. We begin by discussing the prevalence and impact of CVD on Global Health, highlighting the need for effective strategies for the prevention and management of the disease. We then introduce the concept of machine learning and deep learning, explaining the basic principles and different types of

algorithms that are used in these fields. Finally, we outline the research gap that our study aims to fill, explaining how our work contributes to the existing literature on the use of ML and DL for CVD. We conclude this chapter by summarizing the main points and outlining the structure of the rest of the thesis. We will examine the various risk factors for CVD, including genetic and environmental factors, and we will discuss the role of machine learning and deep learning in predicting CVD risk. We will also go over the most recent research on how well these methods predict CVD risk and identify people who are most at risk for the condition. Finally, we will explore the drawbacks and difficulties of applying machine learning and deep learning for the management and prevention of CVD and recommend areas for further investigation in this area. We hope that this discussion will give readers a thorough grasp of how CVD affects global health as well as how machine learning and deep learning could help with disease management and prevention.

The Problems that exist in this domain are given as, Unavailability of efficient detection of cardiovascular disease and having limited treatment options, Incompatible format of available dataset for evaluating the Machine Learning Models and Limited implementation of a standard technique for feature Selection and evaluation of Models. The Objectives of this study have been mentioned as: Select dataset from the online Kaggle Repository and convert it in a format that easily understood by the machine learning models by cleaning and labeling, to improve the predictive performance of the model by identifying and removing irrelevant and redundant features that are not important for predicting cardiac disease, To accurately detect and diagnose Cardiac Disease in an early stage, Evaluate the models by calculating the Accuracy, Precision, Recall and F1-Score of models, to compare Machine Learning-based approaches for performing automated heart disease prediction task.

2 RELATED STUDY

In this section of paper, we give a complete review of the existing literature on the use of ML and DL for the detection and prevention of cardiovascular disease (CVD). We will begin by providing an overview of CVD and its impact on global health. We will then discuss the various ML and DL techniques that have been applied to the study of CVD, including their advantages and limitations. We will also review the existing evidence on the effectiveness of these techniques in predicting CVD risk, diagnosing the disease, and developing personalized treatment plans. Finally, we will identify areas of future research in this field, including potential challenges and opportunities. “Through this review, we aim to provide a comprehensive understanding of the current state of the field and highlight the potential of machine learning and deep learning in improving the prevention” and management of CVD.

Umarani Nagavelli et al. [2] conducted a research study focusing on ML techniques for predicting heart disease. The study provided an overview of various approaches used in detecting heart illnesses. Firstly, the researchers employed a weighted approach with Naive Bayes to predict heart disease. Secondly, an automatic analysis method was proposed, which utilized characteristics from the frequency domain, temporal domain, and information theory to localize and identify ischemic heart disease. SVM and XGBoost classifiers were chosen for the classification in this strategy, as they were found to be the most effective. The third technique involved an improved SVM-based duality optimization methodology for automatically identifying heart failure. To create an efficient “heart disease prediction model (HDPM) for a clinical decision support system (CDSS), the researchers utilized XGBoost for heart disease prediction, a hybrid synthetic minority oversampling technique (SMOTE-ENN) for balancing the training data distribution, and density-based spatial clustering of applications with noise (DBSCAN) for outlier detection and elimination. This comprehensive approach aimed to provide clinicians with a tool to aid in early diagnosis” and improve treatment outcomes.

A research study on CVD was carried out by Shadman Nashif et al. [4] utilising several machine learning models. Through early detection and continued clinical monitoring, the study aimed to lower the mortality rate linked to cardiac conditions. The researchers suggested a WEKA platform-based cloud-based system for the prediction of heart illness that used an efficient machine learning technique chosen after careful consideration of various algorithms. Using 10 times of cross-validation and two well-known open-access databases, the suggested approach was assessed. The accuracy level of the SVM method was 91.53%, while its sensitivity and specificity were 97.50% and 94.94%, respectively. Additionally, an Arduino-based real-time patient monitoring system that could sense variables including body temperature, blood pressure, humidity, and heartbeat was created. The technology allowed carers or medical professionals to continuously monitor patients with heart problems, and it offered real-time sensor information and streaming video in real time for prompt medical assistance. The system also had a feature that would alert the physician through GSM if any real-time medical parameter went above the predetermined limit.

R. Jane Preetha Princy et al. [7] conducted a study focusing on ML models for predicting CVD. With the vast amount of healthcare data generated globally, machine learning algorithms have proven to provide valuable insights when analyzing multidimensional medical datasets. In this study, various supervised “ML algorithms were applied to classify a cardiovascular dataset. The findings indicated that the Decision Tree classification model outperformed other methods such as Naive Bayes, Logistic Regression, Random Forest, SVM, and KNN in predicting cardiovascular illnesses”.

The use of ML techniques in predicting heart disease and cardiovascular disorders shows promise in improving early detection and patient outcomes. These

studies highlight the importance of developing accurate prediction models and utilizing advanced algorithms to handle the complex nature of cardiac conditions. By leveraging machine learning, healthcare professionals can benefit from more precise diagnostic tools and better decision support systems, ultimately leading to improved patient care. The Decision Tree delivered the best outcome with a 73% accuracy rate. This method might help doctors anticipate the onset of heart illnesses and offer the proper treatment in advance. For this investigation, a Kaggle dataset on cardiovascular disease was employed. It has twelve properties, one of which is a target variable. (Table 1) is a representation of the same. For the analysis, people between the ages of 29 and 64 have been considered. Additionally, their weight and height are recorded. The likelihood of developing heart disease increases with one's drinking and smoking habits. Binary values are used to indicate these two variables. The patient is indicated as a "smoker/drinker" if the value is "1," and as a "non-smoker/non-alcoholic" if the value is "0." Patients who engage in regular exercise were indicated with a "1" and others with a "0". The target attribute is the existence or absence of cardiovascular disease. Binary values are included. The "0" stands for normal, whereas the "1" denotes those who have been diagnosed with cardiac disease. Male and female patients were given the gender values 1 and 0, respectively. To calculate the influence, the systolic and diastolic blood pressures are used. The patients' cholesterol and glucose measurements were categorised as normal, above normal, or significantly above normal.

Jilles M Fermont [8] construct a study to predict the CVD by physical Performance measures. The Objectives Although chronic obstructive pulmonary disease (COPD) frequently coexists with CV), it is not yet understood how to more accurately estimate CV risk in those with COPD. Traditional CV risk scores have not been examined specifically in COPD but rather in a variety of populations. Alternative markers may be able to better predict CV risk in people with COPD, although this is unknown. We wanted to know how well traditional CVD risk variables predicted COPD and if any new indicators may enhance prediction beyond standard factors. They used information from the "Global Initiative for Chronic Obstructive Lung Disease (GOLD) stage II–IV COPD cohort: Evaluation of the Role of Inflammation in Chronic Airways Disease, which included 729 patients with stage II–IV COPD. For a median follow-up of 4.6 years, hospital episode statistics and survival data were prospectively gathered". Out of 714 individuals, 192 (27%) required hospitalization for CVD, and 6 passed away as a result. The C-statistic for "the overall CV risk model was 0.689 (95% CI: 0.688 to 0.691). Neither the study outcome nor the model prediction was enhanced by a PWV and CIMT. Independent of traditional risk markers, CRP, fibrinogen, GOLD stage, BODE Index, 4MGS, and 6MWT were all linked to the outcome (p0.05 for all). However, only 6MWT (C=0.727, 95% CI 0.726 to 0.728)" improved model discrimination.

Shaikh Abdul Hannan et al. [9] used a RBF to predict the medical prescription for heart disease in their study. The study included a total of 300 patients' data collected from the Sahara Hospital in Aurangabad. The RBF was used to analyze the patient's data and determine the best medical prescription for the treatment of heart disease. The results of the study showed that the RBF was able to accurately predict the medical prescription for heart disease and was superior to the other methods tested. This study highlights the importance of using the RBF method to accurately predict the medical prescription for cardiac disease and other diseases. It also emphasizes the need for further research in this area to explore the effectiveness of the RBF method in predicting the medical prescription for other diseases. To predict whether a patient would receive a prescription "for heart disease, the radial basis function is used to data on heart disease. The outcomes gained demonstrate that radial basis function can be utilised to successfully prescribe drugs for heart disease. The results point to the importance of accurate diagnosis and the benefits of data training on neural networks-based computerized medical diagnosis systems". Radial basis function neural networks (RBFNs) were implemented in all computation algorithms using M-files, which were based on MATLAB scripts. On a Pentium IV computer with 1 GB of RAM and the Windows XP operating system, all scripts were assembled using the MATLAB compiler. The first step is to gather the patient's medical records and any prescribed medications. The inputs to the data include specifics about the current and past symptoms, and the output is the medications that the doctor has recommended. Step 2 involves converting the signs of heart disease and the medication the doctor has recommended into binary form, which is either 1 or 0. If a symptom or medication is present, it is indicated by a 1, and if neither is present, it is indicated by a 0. Each numeric value is represented by a single digit, either a 0 or a 1. The third step is to train the RBF. To do this, we set the input, hidden, and output layer neuron numbers, learning cycles, and other parameters. Step 4 involves using testing data to gauge how well the trained RBF performed. Medication that the RBF prescribed for the patient with heart disease is step 5.

In the paper by Mai Showman et al. [10], the application of k-NN for the diagnosis of heart disease was discussed. The results showed that compared to a neural network ensemble, k-NN had a higher accuracy rate. However, it was also noted that due to the nature of the data, k-NN was only applicable to narrowly defined categories. This means that while k-NN may be effective in diagnosing a specific type of cardiac disease, it may not apply to more complex cases. Despite this limitation, the authors suggest that k-NN can be a useful tool in the diagnosis of cardiac diseases, especially where the data is limited or highly specific. The authors also suggest that k-NN could be used in conjunction with other methods to improve accuracy. In conclusion, k-NN is an effective tool for diagnosing cardiac

diseases, but its effectiveness is limited to narrowly defined categories. Heart disease has been the top cause of death globally over the past ten years. To help medical professionals identify cardiac illness, researchers have created a number of data mining techniques. One effective data mining method for categorization issues is KNN. The diagnosis of people with cardiac disease, however, uses it less frequently. Recent studies have demonstrated that voting to combine various classifiers outperforms using only one classifier. In this research, the use of KNN to aid medical practitioners in the detection of cardiac disease is examined. It also looks into whether adding voting to KNN can improve how well it diagnoses people with heart disease. The findings demonstrate that using KNN could diagnose heart disease patients with more accuracy than neural network ensemble. The findings also demonstrate that using voting did not improve the KNN's diagnostic accuracy for heart disease.

In a recent study conducted by Jaymin Patel et al. [11], different algorithms of Decision tree classification were compared to determine the best "performance in heart disease diagnosis using WEKA. The algorithms included J48, a logistic model tree, and random forests. The results showed that" the J48 algorithm, which is a type of decision tree, had the highest accuracy, precision, and recall, among the other algorithms. Furthermore, the J48 algorithm was able to identify the cardiac disease with higher accuracy than the other algorithms, making it an ideal choice for heart disease diagnosis. The study concluded that the J48 algorithm was the most suitable choice for heart disease diagnosis in this context and should be used in further analyses. This is an important finding, as it could help increase the precision of diagnosing heart disease and lead to better patient outcomes. Researchers have developed a number of data mining techniques that can assist doctors in identifying heart disease. Less testing might be necessary, though, if data mining techniques are used. An effective and speedy identification strategy is required to lower the number of deaths caused by heart disorders. The decision tree is one of the effective data mining approaches. This study compares different Decision Tree classification algorithms in an effort to improve the efficacy of WEKA in the diagnosis of cardiac disease. The algorithms being tested are the "J48 algorithm, the Logistic model tree method, and the Random Forest algorithm. Existing datasets of individuals with heart illness from the Cleveland database of the UCI repository are utilised to assess and justify the efficacy of decision tree algorithms. This dataset has 303 occurrences and 76 properties. The most promising classification" method will then be suggested for use on a significant amount of data. The aim of this project is to forecast the existence of heart illness in patients, with the presence being classified from unlikely to likely, utilising data mining techniques that are relevant to heart problems.

Abhijeet Jagtap et al. [12] introduced a machine learning study to predict the CVD. In this study they stated that Ineffective analytical tools make it difficult to find hidden links and patterns in CVD dataset. An automated approach for making medical diagnoses might improve the effectiveness of reducing expenses. This website application aims to predict the prevalence of a problem using data from Kaggle and medical studies done by the Cleveland Foundation, notably in the area of heart disease. Data mining techniques are used on the dataset to find hidden patterns that are important to heart ailments in order to forecast the occurrence of heart disease in patients where the presence is assessed on a scale. The prediction of heart disease requires vast volumes of data that are too complex and massive to be gathered and evaluated using conventional approaches. The objective of this paper is to find a ML technique that effectively predicts heart disease and is computationally effective. Data mining is a process that combines statistical analysis, ML, and database technology to uncover hidden patterns and relationships from sizable databases.

Mr. Amol, et al [13] have developed an innovative approach to predicting medical diseases using a combination of support vector machine (SVM) classification and remote sensing (RS). The authors utilized the benefits that RS provides in eliminating redundant information, along with the benefits that SVM provides in identifying complex patterns and classifying data. This method was tested in an Indian medical school and resulted in a higher accuracy rate than existing methods in predicting cardiac diseases. The approach is promising in that it could be applied to a variety of otherwise difficult-to-diagnose medical conditions. Currently, the authors are working to refine the approach even further and apply it to other medical diseases as well. Ultimately, this approach may revolutionize the way medical conditions are diagnosed, providing doctors with more accurate and timely information. In their paper, Sarath Babu et al [21] compares classification methods by bagging algorithm to predict heart disease. Their study showed that the bagging algorithm was superior to other methods in terms of both performance rate and accuracy level. This method works by taking multiple samples from a dataset and then training them on different combinations of features and data points. The bagging algorithm then evaluates the results, and the model with the highest accuracy is chosen as the best. This allows for better accuracy than other methods and improved performance. In their study, the bagging algorithm proved to be effective in predicting cardiac disease with a high level of accuracy. This is great news for medical professionals, as it provides a reliable and accurate way of predicting heart disease. Furthermore, the bagging algorithm can be used for other classification tasks, such as predicting cancer outcomes or predicting stock market movements. Therefore, this study shows the potential of the bagging algorithm and its effectiveness in solving a variety of classification tasks. O.E. Taylor et al [22] presented in their study compared the performance of KNN, SVM, Random classifier, and Decision Tree classifier for the Heart Disease Prediction System. The Decision Tree method produced the most accurate result, with a prediction accuracy of 98.83%, outperforming the other methods.

Table 1. Techniques used in literature

Ref	Year	Method	Algorithems	Classes	Dataset Size	Tool
Umarani Nagavelli et al. [2]	2022	ML	Naïve Bays, Support Vector Machine and XGBOOST	Binary	297 records	Python
Shadman Nashif et al. [4]	2018	ML	SVM, RF, NB, ANN,	Binary	303 Record	Weka
R. Jane Preetha Princy et al. [7]	2020	Supervised ML	SVM, RF, NB, DT, KNN, LR	Binary	70000 Records	Python
Jilles M Fermont [8]	2020	Multicentre Observational	COPD	Multi	729 Records	Manually
Shaikh Abdul Hannan et al. [9]	2019	Deep Learning	RBF	Binary	300 Records	MATLAB
Mai Showman et al. [10]	2012	ML	KNN	Binary	303 Records	Python
Jaymin Patel et al. [11]	2016	ML	RF, LR, andJ48	Binary	303 Records	Weka
Abhijeet Jagtap et al. [12]	2019	ML	SVM, NB and LR	Binary	303 Record	Weka
Mr. Amol, et al [13]	2019	Deep Learning	RBF	Binary	300 Records	MATLAB
Proposed	2023	Hybrid	SVM, XGBOOST, NB, KNN, DT	Binary	70001 Records	Google Colab

Table 2. Accuracy and confusion matrix of literature studies

Ref	Features	Accuracy	Presicion	Recall	F1-Score
Umarani Nagavelli et al. [2]	temperature, blood pressure, humidity, heartbeat etc.	94.03 %	82.34 %	81.3 %	89.21 %
Shadman Nashif et al. [4]	blood pressure, humidity, heartbeat etc.	91.53 %	N/A	N/A	N/A
R. Jane Preetha Princy et al. [7]	Age, Height, Width, Gender etc.	73 %	75 %	78 %	73 %
Jilles M Fermont [8]	Different Symptoms	65 %	N/A	N/A	N/A
Shaikh Abdul Hannan et al. [9]	Previous, Present, Personnel History, Physical Examination, Cardio Vascular System, Respiratory Rate etc.	75 %	N/A	N/A	N/A
Mai Showman et al. [10]	Age, Height, Width, Gender etc.	91 %	N/A	N/A	N/A
Jaymin Patel et al. [11]	age, gender, resting blood pressure, cholesterol, fasting blood, age, gender	55.4 %	N/A	N/A	N/A
Abhijeet Jagtap et al. [12]	age, gender, resting blood pressure, cholesterol, fasting blood, age, gender	64.5 %	N/A	N/A	N/A
Mr. Amol, et al [13]	Age, Height, Width etc.	89 %	N/A	N/A	N/A
Proposed	age, gender, resting blood pressure, cholesterol, fasting blood, age, gender	92.34	92%	85%	86%

3 PROPOSED METHOD

Our Proposed Framework contain four phases that is depicted in figure 1 In our proposed methodology. In first phase, we Select a suitable Cardiovascular dataset from the Kaggle repository and then clean the dataset by removing the missing values and then labeled the categorical data into numeric values to evaluate on the ML models. In second phase, Split the dataset by using two techniques (Traing&Testing split and Cross-Validation) for the training and Testing of models. In third phase, we apply five different Machine Learning models (RF, KNN, DT, NB and XGBOOST) on the processed dataset. Finally evaluate the ML models on the basis of Training accuracy, testing accuracy, Presicion Recall and F1-Score and compare their results.

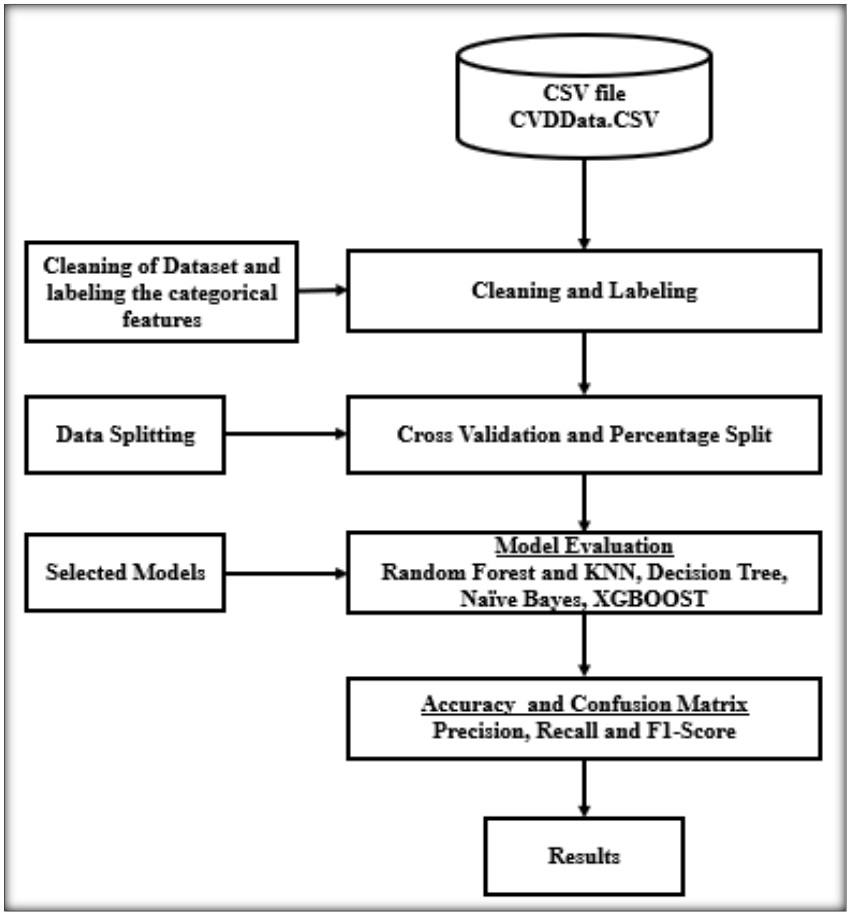

Figure 1. Proposed Method

3.1 Phase_1

In this Cardio Vascular Disease dataset is downloaded from the Kaggle ("https://www.kaggle.com/datasets/bhadaneeraj/cardio-vascular-disease-detection"). The dataset contains 12 features that displayed in the table 4. The second part of this phase is to pre-process the by converting string values into numeric and filling the blank spaces in the csv file.

3.2 Phase_2

The second phase of our methodology is splitting of dataset into training and testing of models. We use two methods of data splitting (Training Testing Split and Cross_Validation). In Training Testing Split 80 percent of dataset is used for training of models and remaining 20 percent dataset is used for testing of selected models. In Cross_Validation we use 10K cross Validation.

Table 3. Splitting of dataset

Parameters	Percentage	Number of records
Training Dataset	80	56001
Testing Dataset	20	14,000
Total	100	70001

Table 3 shows that our selected dataset contains Seventy thousand record in which fourteen thousand records

are used for Training of models and Fifty-six thousand records are used for Testing of models.

3.3 Phase_3

The third phase of our proposed methodology is selection of machine learning models. We select five machine learning classifiers for the evaluation of our cleaning dataset. such as Decision Tree (DT), K-Nearest Neighbors (KNN), Naïve Bayes (NB), Logistic Regression (LR) and Random Forest.

4 RESULTS

This section includes a review of the experimental findings as well as the performance measures and our suggested method for performance evaluation and results. We use python language to build machine models by using google Colab platform.

The used system having Hard disk drive of 320 Giga Byte and Solid-State Disk of 256 Giga Byte and Random-Access Memory of 8 Giga Byte. Windows 10_Pro operating system is installed on the system. The system is Core i5 4th generation and dell latitude E6440 architecture. Different tools are used for writing of thesis and implementation of machine learning models.

4.1 Evaluation of Models

We evaluate our proposed model on following two performance measures:

- Training Accuracy
- Testing Accuracy

Each of the performance measures are calculated for all five selected machine learning models.

The analysis of result in table 4 shows that model DT and RF is well perform on the training dataset with accuracy of 99. 99 percent and 99.98 percent and model NB, RF and XGBOOST are performs good on the testing dataset with accuracy of 92.34 percent, 92.06 percent and 92.31 percent.

Table 4. Comparison of Accuracy

Models	Training Accuracy (%)	Testing Accuracy (%)
DT	99.99	85.46
RF	99.98	92.06
NB	92.51	92.34
KNN	78.76	71.48
XGBOOST	92.73	92.31

Figure 2 shows the accuracy graph of models. X-Axis contains the name of each models with training and testing labels and Y-Axis shows the accuracy value of the models ranges from 0 to 100 percent.

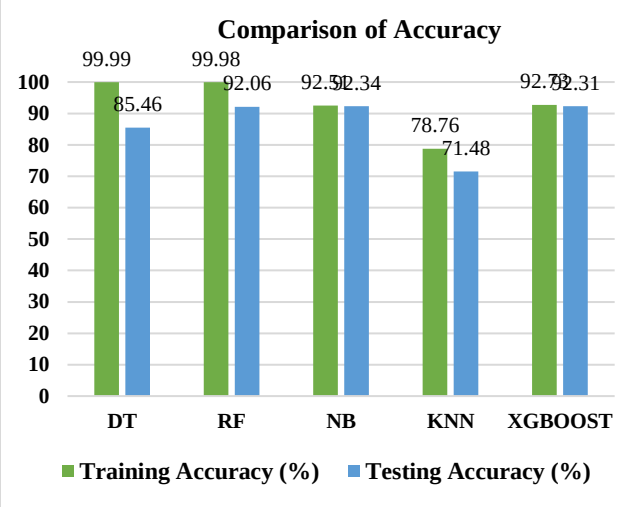

Figure 2 Comparison Graph

Table 5 Dataset Example

	age	gender	Height	weight	ap_hi	ap_lo	cholesterol	gluc	smoke	alco	active	Cardio
0	18393	2	168	62.0	110	80	1	1	0	0	1	0
1	20228	1	156	85.0	140	90	3	1	0	0	1	1
2	18857	1	165	64.0	130	70	3	1	0	0	0	1
3	17623	2	169	82.0	150	100	1	1	0	0	1	1
4	17474	1	156	56.0	100	60	1	1	0	0	0	0
5	21914	1	151	67.0	120	80	2	2	0	0	0	0
6	22113	1	157	93.0	130	80	3	1	0	0	1	0
7	22584	2	178	95.0	130	90	3	3	0	0	1	1
8	17668	1	158	71.0	110	70	1	1	0	0	1	0
9	19834	1	164	68.0	110	60	1	1	0	0	0	0

5 CONCLUSION

Nearly 19 million people died each year from cardiovascular and chronic respiratory diseases, which were a global threat. It was necessary to address the causes of these diseases because of the high death rate. The investigation uncovered a number of causes, but the inability to forecast these diseases symptoms was by far the most significant. In this work, we developed a method for anticipating these diseases crucial symptoms, which would aid in early disease diagnosis and allow patients to begin treatment. This research introduced a new computational medicine research using machine learning (ML) paradigms to forecast cardiovascular disease (CVD). Data were processed by methods in sequence with various parameters. Different models were created that predicted CVD risk based on individual age, gender, ethnicity, body mass, etc., and lifestyle factors. The research also focused on performing a complete comparison of ML models. We applied five ML-based algorithms such as Decision Tree (DT), K-Nearest Neighbors (KNN), Naïve Bayes (NB), XGBOOST, and Random Forest and evaluated these models on the basis of Training and Testing and also calculated the Precision Recall and F1-Score for each model. Naïve Bayes and XGBOOST Classifier performed better with an accuracy of 92.31 and 92.34 percent as compared to other models.

6 FUTURE WORK

In future this work can be extended by using different ensemble machine learning models to improve the accuracy and also use deep learning techniques to fine tune the dataset. The classes of the dataset can be extended in the future.

CREDIT AUTHOR STATEMENT

"Fouzia Kanwal: Conceptualization, Methodology, Software **Mr. Kamran Abid:** Supervisor **Muhammad Sajid Maqbool**: Data curation, Writing-Original draft preparation, Visualization, Investigation **Naeem Aslam:** Review **Muhammad Fuzail:** Data creation and software working"

COMPLIANCE WITH ETHICAL STANDARDS

"It is declared that all authors don't have any conflict of interest. Furthermore, informed consent was obtained from all individual participants included in the study".

REFERENCES

[1]. K. Reynolds, A. S. Go, T. K. Leong, D. M. Boudreau, A. E. Cassidy-Bushrow, S. P. Fortmann, et al., "Trends in incidence of hospitalized acute myocardial infarction in the Cardiovascular Research Network (CVRN)," The American Journal of Medicine, vol. 130, no. 3, pp. 317-327, 2017.

[2]. U. Nagavelli, D. Samanta, and P. Chakraborty, "Machine Learning Technology-Based Heart Disease Detection Models," Journal of Healthcare Engineering, vol. 2022, pp. 1-9, 2022. doi: 10.1155/2022/7351061.

[3]. B. Troesch, H. K. Biesalski, R. Bos, E. Buskens, P. C. Calder, W. H. Saris, et al., "Increased intake of foods with high nutrient density can help to break the intergenerational cycle of malnutrition and obesity," Nutrients, vol. 7, no. 7, pp. 6016-6037, 2015.

[4]. S. Nashif, M. R. Raihan, M. R. Islam, and M. H. Imam, "Heart disease detection by using machine learning algorithms and a real-time cardiovascular health monitoring system," World Journal of Engineering and Technology, vol. 6, no. 4, pp. 854-873, 2018.

[5]. L. C. Mantoani, S. Dell'Era, W. MacNee, and R. A. Rabinovich, "Physical activity in patients with COPD: the impact of comorbidities," Expert Review of Respiratory Medicine, vol. 11, no. 9, pp. 685-698, 2017.

[6]. M. Heron, "Deaths: Leading causes for (2010)," Natl. Vital Stat. Reports, vol. 62, no. 6.

[7]. R. J. P. Princy, S. Parthasarathy, P. S. Hency Jose, A. Raj Lakshminarayanan, and S. Jeganathan, "Prediction of Cardiac Disease using Supervised Machine Learning Algorithms," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), 2020, pp. 570-575, doi: 10.1109/ICICCS48265.2020.9121169.

[8]. J. M. Fermont, M. Fisk, C. E. Bolton, W. MacNee, J. R. Cockcroft, J. Fuld, et al., "Cardiovascular risk prediction using physical performance measures in COPD: results from a multicentre observational study," BMJ Open, vol. 10, no. 12, p. e038360, 2020.

[9]. A. Shaikh Abdul Hannan, A. V. Mane, R. R. Manza, and R. J. Ramteke, "Prediction of Heart Disease Medical Prescription using Radial Basis Function," IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Dec. 2010, pp. 28-29, doi: 10.1109/ICCIC.2010.5705900.

[10]. M. Shouman, T. Turner, and R. Stocker, "Applying k-Nearest Neighbors in Diagnosing Heart Disease Patients," International Journal of Information and Education Technology, vol. 2, no. 3, pp. 220-223, Jun. 2012.

[11]. J. Patel, T. Upadhyay, and S. Patel, "Heart Disease Prediction using Machine Learning and Data Mining Technique," vol. 7, no. 1, pp. 129-137, Sep. 2015-Mar. 2016.

[12]. A. Jagtap, P. Malewadkar, O. Baswat, and H. Rambade, "Heart disease prediction using machine learning," International Journal of Research in Engineering, Science and Management, vol. 2, no. 2, pp. 352-355, 2019.

[13]. W. A. A. Wghmode, D. Sawant, and D. D. Ketkar, "Heart Disease Prediction Using Data Mining Techniques," Heart Disease, 2017.

[14]. M. Anbarasi, E. Anupriya, and N. C. S. N. Iyengar, "Enhanced prediction of heart disease with feature

subset selection using genetic algorithm," International Journal of Engineering Science and Technology, vol. 2, no. 10, pp. 5370-5376, 2010.

[15]. E. Taylor, P. S. Ezekiel, and F. B. Deedam-Okuchaba, "A Model to Detect Heart Disease using Machine Learning Algorithm," International Journal of Computer Sciences and Engineering, 2019.

[16]. M. Singh, L. M. Martins, P. Joanis, and V. K. Mago, "Building a Cardiovascular Disease Predictive Model using Structural Equation Model & Fuzzy Cognitive Map," IEEE International Conference on Fuzzy Systems (FUZZ), 2016, pp. 1377-1382.

[17]. M. Gudadhe, K. Wankhade, and S. Dongre, "Decision Support System for Heart Disease Based on Support Vector Machine and Artificial Neural Network," International Conference on Computer and Communication Technology (ICCCT), Sept. 2010, pp. 17-19, doi: 10.1109/ICCCT.2010.5640377.

[18]. M. L. Sharan and B. S. Kumar, "Analysis of Cardiovascular Heart Disease Prediction Using Data Mining Techniques," International Journal of Modern Computer Science, vol. 4, no. 1, pp. 55-58, Feb. 2016.

[19]. D. S. Medhekar, M. P. Bote, and S. D. Deshmukh, "Heart Disease Prediction System Using Naive Bayes," International Journal of Enhanced Research in Science Technology & Engineering, vol. 2, no. 3, pp. 1-5, Mar. 2013.

[20]. N. Ajam, "Heart Diseases Diagnoses Using Artificial Neural Network," Network And Complex Systems, vol. 5, no. 4, pp. 7-11, Feb. 2015. ISSN: 2224-610X (Paper), ISSN: 2225-0603 (Online).

[21]. B. Bahrami and M. H. Shirvani, "Prediction and Diagnosis of Heart Disease by Data Mining Techniques," Journal of Multidisciplinary Engineering Science and Technology (JMEST), vol. 2, no. 2, pp. 164-168, Feb. 2015. ISSN: 3159-0040.

[22]. M. S. Maqbool, I. Hanif, S. Iqbal, A. Basit, and A. Shabbir, "Optimized Feature Extraction and Cross-Lingual Text Reuse Detection using Ensemble Machine Learning Models," Journal of Computing & Biomedical Informatics, vol. 5, no. 01, pp. 26-40, 2023.

[23]. U. Fazal, M. Khan, M. S. Maqbool, H. Bibi, and R. Nazeer, "Sentiment Analysis of Omicron Tweets by using Machine Learning Models," Journal of Computing & Biomedical Informatics, 5(01), pp. 82-95, 2023.

[24]. M. A. Hasnain, S. Ali, H. Malik, M. Irfan, and M. S. Maqbool, "Deep Learning-Based Classification of Dental Disease Using X-Rays," Journal of Computing & Biomedical Informatics, vol. 5, no. 01, pp. 82-95, 2023.

[25]. A. Basit, I. Hanif, M. S. Maqbool, W. Qayyum, M. A. Hasnain, and R. Nazeer, "Cross-Lingual Information Retrieval in a Hybrid Query Model for Optimality," Journal of Computing & Biomedical Informatics, vol. 5, no. 01, pp. 130-141, 2023.

[26]. V. Shorewala, "Early detection of coronary heart disease using ensemble techniques," Informatics Med. Unlocked, vol. 26, p. 100655, Jan. 2021, doi: 10.1016/J.IMU.2021.100655.

[27]. I. Gupta, R. Shangle, V. Latiyan, and U. Soni, "Cardiovascular Disease Detection using Artificial Immune System and other Machine Learning Models," J. Phys. Conf. Ser., vol. 1950, no. 1, p. 012032, Aug. 2021, doi: 10.1088/1742-6596/1950/1/012032.

[28]. M. I. H. Khan and M. R. H. Mondal, "Data-Driven Diagnosis of Heart Disease," Int. J. Comput. Appl., vol. 176, no. 41, pp. 46-54, Jul. 2020, doi: 10.5120/IJCA2020920549.

[29]. S. Palaniappan and R. Awang, "Intelligent heart disease prediction system using data mining techniques," in 2008 IEEE/ACS International Conference on Computer Systems and Applications, 2008, pp. 108-115.

[30]. A. Taneja, "Heart disease prediction system using data mining techniques," Oriental Journal of Computer Science and Technology, vol. 6, no. 4, pp. 457-466, 2013.

[31]. S. Anitha and N. Sridevi, "Heart disease prediction using data mining techniques," Journal of Analysis and Computation, vol. 6, no. 4, pp. 457-466, 2019.

[32]. C. S. M. Wu, M. Badshah, and V. Bhagwat, "Heart disease prediction using data mining techniques," in Proceedings of the 2019 2nd International Conference on Data Science and Information Technology, 2019, pp. 7-11.

[33]. C. Krittanawong, H. Zhang, Z. Wang, M. Aydar, and T. Kitai, "Artificial intelligence in precision cardiovascular medicine," Journal of the American College of Cardiology, vol. 69, no. 21, pp. 2657-2664, 2017.