A Comprehensive Study on Early Prevention and Detection of Cardiac Health Issues Using Machine Learning and Deep Learning Algorithms

Shah Hussain Bangash[1], Irfan Ullah Khan[1], Zabi Ullah Khan[2], Junaid Khan[1], Mohsin Tahir[1], Waqas Ahmad[1], Ihtisham Ul Haq[3], Maimoona Asad[4], Mohammad Mansoor Qadir[2]

[1]Iqra National University, Peshawar, Pakistan.
[2]CECOS University of IT and Emerging Sciences, Peshawar, Pakistan
[3]Department of ICT, University of Calabria, Italy
[4]Department of Information and Communication Engineering from the College of Electronic and Information Engineering, Shenzhen University, China

***Abstract*:** Heart disease is a major global health concern and a primary cause of death and morbidity, affecting millions of people worldwide. Machine learning techniques, such as data mining tools, have been applied to improve the performance of medical diagnosis for heart disease. The complexity of heart disease calls for a thorough investigation of its various facets, including coronary artery disease, heart failure, arrhythmias, valvular heart disorders, and congenital heart defects. Risk factors for heart disease include genetics, environmental influences, nutrition, exercise, smoking, excessive alcohol use, obesity, diabetes, and hypertension. Early detection and risk assessment are made possible by advanced imaging techniques and diagnostic technologies, leading to timely interventions and improved cardiac health. In this work, we proposed both Machine learning algorithms including Support Vector Machine (SVM), Decision Tree, KNN, XGB classifier, and Logistic Regression and Deep learning algorithm based on the CNN model to show some useful results to improve the diagnostics related to cardiac issues. The study aims to explore the prevalence, risk factors, diagnostic techniques, therapeutic approaches, and efforts to lessen the negative effects of heart disease on public health.

Index Terms- Machine Learning Techniques; Deep Learning Algorithms; Detection; Heart disease classification; Performance Measurements; Feature Selection

I. INTRODUCTION

Over the last two decades, researchers and scientists have struggled to control heart disease issues for human life. However, 30 million people suffer from heart disease due to smoking, high blood pressure, high cholesterol, poor hygiene, viral infection, unhealthy diet, and other diseases [1-2]. The machine learning techniques classification of heart disease applied data mining tools improving performance measures of the medical diagnosis. Data mining has many tools, but some are impressive, like Kanime, Orange, MATLAB, RapidMiner, and Scikit-Learn techniques [3-4-5]. The early detection of disease has more chances to protect and reduce the death toll around the globe [6]. Cardiovascular disease, sometimes referred to as heart disease, is a major global health concern and a primary cause of death and morbidity [7]. It includes a variety of illnesses that affect the heart and blood arteries, preventing them from operating normally and perhaps resulting in serious problems [8]. The frequency of heart disease has increased as cultures adopt more sedentary lifestyles and nutritional changes, which has sparked intense research efforts to comprehend its underlying causes, risk factors, and viable therapies [9-10]. The complexity of cardiac disease calls for a thorough investigation of all its facets. One of the most prevalent types of heart disease is coronary artery disease, which is defined by the accumulation of plaque in the coronary arteries [11-12]. The heart muscle's blood flow is constrained because of this disorder, which can cause angina, heart attacks, and possibly fatal arrhythmias [13]. Heart failure, arrhythmias, valvular heart disorders, and congenital heart defects are further forms of heart illness. Each presents unique difficulties and requires a different strategy for therapy. Heart disease has a complicated etiology that frequently combines genetic predisposition with environmental influences. Heart disease risk factors include nutrition, exercise, smoking, and excessive alcohol use. All these factors have a major impact [14-15-16]. Furthermore, underlying diseases including obesity, diabetes, and hypertension raise the risk of cardiovascular problems. Heart disease research has provided critical insights into the pathophysiology of the condition and available therapies. Early detection and risk assessment are made possible by cutting-edge imaging techniques and diagnostic technologies, allowing healthcare professionals to act before irreversible harm is done [17-18]. The creation of drugs that efficiently manage risk factors, treat symptoms, and improve overall cardiac health is the result of advances in pharmaceuticals. Machine learning techniques provide opportunities to predict CHD and remove the complexities of the data and correlation challenges [19-20]. The probabilistic model's performance measures of evaluation empirical results through Naïve Bayes promise to detect CHD. With the help of technology, classifying massive amounts of data generated in hospitals and medical institutes becomes electronic format [21]. The data drastically enhanced in the field of medical history screening patients' detection health system. Preventing heart disease is valuable to check on accurate time safe from failure and early treatment. In the proposed study, machine learning classifies reliable non-invasive methods efficient for healthcare system diagnosis. The classifier's performance pre-processing to validate the system also struggles to reduce the impact of algorithms based on accuracy and improve the execution time efficiently [22-23]. Heart disease (HD) is considered a complex, challenging task to protect the deadliest and provide the required blood for a body. The symptoms of HD, feeling weakness in the body, fatigue, breath shortness, and swollen feet, are signs of positive functional cardiac. HD disease denotes alarming situations, especially in developing countries, using all the resources to find accurate treatment reducing risks to human life [24-25]. The care of heart disease has also changed because of interventions like cardiac rehabilitation programs, lifestyle changes, and surgical procedures like angioplasty and bypass surgery, which have increased both the quality of life and survival rates for those who are affected [26]. This study will explore the complex web of heart disease, including its prevalence, risk factors, diagnostic techniques, therapeutic approaches, and continuous efforts to lessen its negative effects on public health. By thoroughly comprehending the complex nature of heart disease, we may seek to create more focused and efficient interventions to lessen its worldwide impact and improve cardiovascular health for everyone [27-28].

II. 2. Literature Review

Heart disease is a severe global issue affecting the heart in terms of various problems: heart failure, blood vessel disease, arrhythmia, heart attack, stroke, etc. Ilias Tougui et al. proposed comparing data mining tools and classifying data through machine learning techniques to predict a complex workflow design process. When the author compared the data mining tools' performance of accuracy found out each module has strengths and weaknesses, while Knime is better than others in terms of usability [29]. R. Kannan et al. studied heart disease diagnosing with the help of machine learning classifiers Roc Curve. The researchers aim clearly to determine and compare machine learning experimental analysis to predict the ROC cure morality heart disease creating between 30 to 50 age groups. Medical field analysis is one of the best methods to collect valuable information about the issues facing our healthcare system [30].FadiThabtah et al. highlight some critical machine learning techniques to measure the accurate prediction of medical data analysis and distribute the altogether better associated. Machine learning can use intelligent methods to process, explore, and interpret training models to deal with the facing problem [31]. Amin UlHaq et al. proposed a system to reduce the computation time of classifiers and improve accuracy through a hybrid framework using machine learning algorithms. The selected features performance was underestimated in terms of LASSO, mRMR, and some classification classifiers evaluation metrics computed [32]. Rohit Bharti et al. worked on combined Deep learning and Machine learning techniques to predict cardiac arrhythmia dealing with the high dimensionality of medical data. The optimization approaches increased the evaluation results and trained the models to improve the accuracy [33].

In the below table, SVM stands for support vector machine, LG stands for Liner Regression, KNN K-Nearest Neighbors, NB stands for Naïve Bayes, RF stands for Random Forest, DL stands for Deep Learning, HRFLM Stands for Hybrid Random Forest with Linear Model

Table #1 Previous Study of Heart Disease

Authors	Algorithms	Purpose	Contribution	Weakness Limitation
[1]	LG, SVM, K Nearest Neighbors, ANN, NB, and RF	Compare six common data mining tools: Orange, Weka, RapidMiner, Knime, MATLAB, and Scikit-Learn	Improving the medical diagnosis	Complex workflow design
[2]	LG, RF, Stochastic gradient boosting, SVM	Compared the best model of each machine learning algorithm with the ROC curve	Improving the health of a patient	Deep learning tensor flow will automate and increase the process of prediction in terms of speed
[3]	Naïve Bayes (NB), Support Vector Machine (SVM) and Decision Tree (DT)	The most effective ML models	Further need Accuracy,	complexity of the data and correlations
[4]	RELIEF Algorithm, mRMR Algorithm, Logistic regression, K-nearest neighbor, SVM, Naive Bayes	Feature selection	Improve the performance of classifiers	Reduced the computation time, irrelevant features reduced the performance, reduced the execution time
[5]	Integrate heart-disease-trained ML and DL models, Gaussian distribution	ML algorithms and DL are applied to compare the results and analysis of the UCI Heart Disease dataset.	Getting better results	The dataset is not that large
[6]	Hybrid Random Forest with Linear Model (HRFLM), Decision tree, NB, GLM, SVM, LR, DT	Performance comparison with various models	improving the accuracy	Diverse mixtures of machine learning techniques to better prediction techniques
[7]	Support vector machine, Logistic regression, Artificial neural network, K-nearest neighbor, Naïve bays, and Decision tree	Feature Extraction, Feature Selection and classification	Achieved good accuracy as compared to previously proposed methods	feature Fasting blood sugar (FBS) is not a suitable heart disease diagnosis. Reduce the execution time.
[8]	Naïve Bayes, Decision list and k-NN	Feature Extraction, Feature Selection, classification, Regression	Enhanced accuracy than an unclean one with missing values	Making models that can predict whether a patient is likely to develop heart disease or not.
[9]	Support Vector Machines (SVM), K-Nearest Neighbor (KNN), Bayes, Decision, Trees (DT), Random Forest (RF), and	Feature Extraction, Feature Selection, classification, Regression	Achieve an accuracy of 94.12%	Decision trees have performed very poorly in some other cases which could be due to overfitting.

	ensemble models			
[10]	Naive Bayes, Decision Tree, Logistic Regression, and Random Forest	Comparative study by analyzing the performance of different machine learning algorithms	Compares the accuracy score	Enhanced by developing a web application based on the Random Forest algorithm as well as using a larger dataset as compared
[11]	K-nearest neighbor, decision tree, linear regression, and support vector machine (SVM)	Analysis of the performance of algorithms	Calculate the accuracy of machine learning algorithms	Low Accuracy
[12]	Artificial neural network (ANN), decision tree (DT), random forest (RF), support vector machine (SVM), naïve Bayes) (NB) and k- nearest neighbor algorithm	Survey type, Feature Selection	To avoid these errors and to achieve better and faster results	Better to use search algorithms for selecting the features and then applying ML techniques for prediction will give us better results in the prediction of heart disease.
[13]	Decision Tree Algorithm and Naive Bayes Algorithm	Prediction, Classification techniques	Analyze the best algorithm in terms of accuracy	Extended or improved for the automation of heart disease analysis including some other machine learning algorithms.
[14]	Neural network, K-nearest neighbor, naive Bayes, and logistic regression	Detection,	Model which contributes to increasing the overall accuracy of the algorithm	In the future models of cardiac disease detection system
[15]	Decision Tree (DT), Naïve Bayes (NB), Multilayer Perceptron (MLP), K-Nearest Neighbour (K-NN), Single Conjunctive Rule Learner (SCRL), Radial Basis Function (RBF) and Support Vector Machine (SVM)	Bagging, boosting, and stacking approaches	Compare the different machine learning techniques on a small dataset to improve the highest accuracy	Accepted Limited dataset instance
[16]	Logistic Reg,Naïve Bayes, SVM KNN, Decision Tree, Random Forest ANN, DNN, MLP	Comparative Study and Analysis	Best results in terms of accuracy and other evaluating metrics. Fast, accurate, less computation time, Good in Accuracy, Fast, very less computation power	High computation power consumption, less accurate, High computation power consumption, very slow for large datasets, very less accurate takes a lot of secondary
				memory for data storage, Slow for large datasets, high computation power
[17]	Naïve Bayes, decision tree, K-nearest neighbor, and random forest algorithm.	Comparative result of classification techniques	Highest accuracy score is achieved	Complex and combination of models to get higher accuracy
[18]	k-nearest neighbor (KNN), decision tree (DT), and random forests (RF) al-algorithms the RF, MLP, and KNN	Classification, Prediction, detection	Identify machine learning classifiers with the highest accuracy for such diagnostic	Some algorithm's accuracy does not have good results
[19]	SVM, KNN, ANN, CNN, GAN	systematic literature review (SLR) approach imbalanced data in heart disease predictions	Dealing with imbalanced data	ML-based heart disease diagnosis with imbalanced data still has unexplored aspects and many potentials to unlock in the coming years.
[20]	K nearest neighbors, Novel KNN approach,	Classification and validation	Accuracy is improved as well as classification and validation	Increase still accuracy in some percentage
[21]	Logistic regression, KNN, naive Bayes, Random Forest Classifier	Computer-aided techniques	Improve the accuracy model in finding the probability of the classifier to correctly and accurately identify heart disease	Low accuracy of algorithms
[22]	Hybrid Random Forest with a Linear Model (HRFLM), SVM, RF, Decision Tree	Comparative study	Highest accuracy when compared with other models	Diverse mixtures of machine learning techniques to better prediction techniques
[23]	K-Nearest Neighbours Classifier, Support Vector Classifier, Decision Tree Classifier, and Random Forest Classifier	Model Detection, Feature Selection	Extended to a real-time system using a Deep Learning approach, where users can upload their test results as image	Machine Learning is best while working on deep learning techniques
[24]	HMM, wearable sensors	Feature Selection	The resulting model can achieve good accuracy	Larger frameworks for remotely monitoring a patient's health state in a clinically meaningful manner

[25]	Artificial Neural Network (ANN), Decision tree, Fuzzy Logic, K-Nearest Neighbour (KNN), Naïve Bayes and Support Vector Machine (SVM)	Data Analytics Approach, Feature Selection	feature selection methods to improve the accurate performance of algorithms	testing different discretization techniques, multiple classifiers voting techniques and different decision tree types information gain and gain ratio
[26]	Support Vector Machine, Decision Tree, Naïve Bayes, K-Nearest Neighbour, and Artificial Neural Network.	Survey Type Paper, Data Mining Techniques	Improve the accurate performance of algorithms	poor results

2.1. Heart Diseases

According to the current situation of heart disease patients almost 30 million people are suffering throughout the world. It is life-frightening for the affected patients experiencing the stress of living contemplating global health precedence [34]. The researchers and experts in healthcare highlight the main reasons the heart disease patients various reasons, such as unhealthy diet, high blood sugar, physical inactive, smoking, viral infection, blood pressure, poor hygiene, cholesterol, etc. The heart disease patient should remember the signs and symptoms of heart disease caused a deep investigation of the current major reasons of the following highlighted. Heart disease, commonly referred to as cardiovascular disease, continues to be a major problem for world health, as it contributes significantly to morbidity, mortality, and healthcare expenditures [35-36]. The goal of this research article is to present a thorough summary of current discoveries in the knowledge and treatment of cardiac disease, emphasizing significant advancements in diagnosis, therapy, and prevention [37].

III Research Methodology

Heart disease prevention is still a top priority, and recent research emphasizes a multifaceted strategy. Reduced cardiovascular risk is largely attained by lifestyle changes such as increased physical activity, good eating, and quitting smoking. The development of wearable equipment and smartphone applications that allow for real-time monitoring of heart health has been made possible by advancements in digital health technology, empowering people to take preventative action. Global healthcare systems continue to be challenged by heart disease, but recent developments in diagnosis, therapy, prevention, and precision medicine provide new hope for patient outcomes.

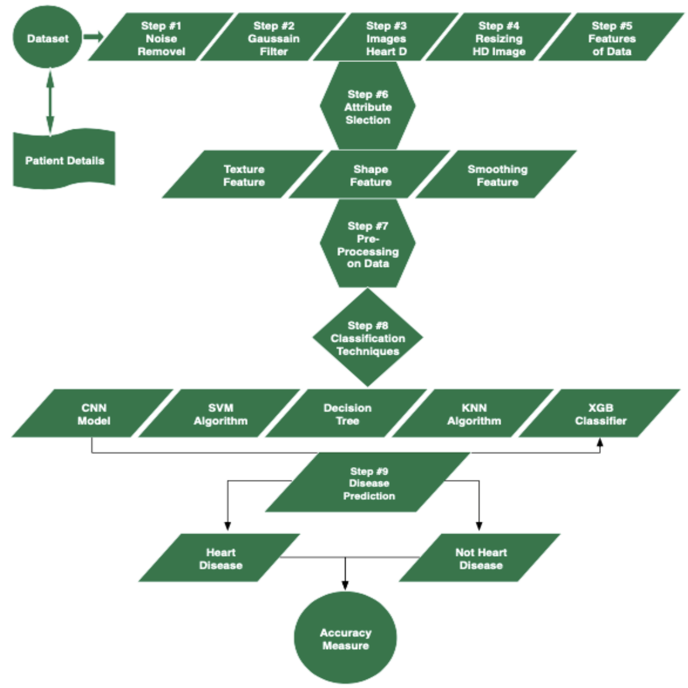

Figure #1: Block Diagram of Processing

Figure 1 shows machine learning is a technique that forecasts the potential development of heart disease. Additionally, machine learning is utilized to solve several problems. The prediction of a result based on previously collected data is the fundamental application of machine learning. To predict the outcome, the machine applies the designs from the present dataset to a mysterious dataset. In AI, the order method is typically applied to expectations. The block diagram explained the complete process of analyzing heart disease images through machine and deep learning algorithms. It's a complex process to analyze and determine the performance of medical images to achieve the highest accuracy.

In recent years, improvements in diagnostic equipment have completely changed how early heart disease identification and evaluation are conducted. Cardiovascular magnetic resonance imaging (MRI) and computed tomography angiography (CTA) are non-invasive imaging methods that allow for thorough viewing of the structure and operation of the heart [38-39]. Additionally, the development of biomarkers like natriuretic peptides and high-sensitivity troponin enables more precise risk classification and prognosis prediction.

Heart disease treatment has advanced significantly, enabling a variety of therapeutic choices to meet the needs of each patient [40]. New anticoagulants and antiplatelet medicines have improved effectiveness and safety profiles thanks to pharmacological treatments. Invasive techniques including transcatheter aortic valve replacement (TAVR) and percutaneous coronary intervention (PCI) have improved in sophistication, lowering risks and improving patient outcomes [41]. Additionally, gene editing methods and regenerative medicines show promise for treating the underlying causes of some genetic heart illnesses as well as restoring damaged heart tissue.

Healthcare practitioners may improve tactics to battle this ubiquitous disease and improve the quality of life for affected people by staying at the forefront of research and innovation [42-43]. To advance these revolutionary developments in the field of cardiac disease, continued cooperation between researchers, physicians, and policymakers is imperative. These are the main causes of heart diseases of the following [44-45].

Shortness of breath: The effects of cardiac disease on the heart's operation and blood flow may result in insufficient oxygen delivery to the body. As a result, the body struggles to get enough oxygen into the bloodstream, resulting in dyspnea during exercise or even while at rest.

Chest pain: The sensation of pressure or discomfort in the chest may be a sign of several underlying problems. It could be a sign of a cardiac condition like angina or a heart attack. As chest symptoms may point to potentially serious health issues, prompt medical diagnosis is essential to ascertain the cause and guarantee prompt treatment.

Chest Pressure: A feeling of tightness or weight in the chest region is referred to as chest pressure. It may be linked to heart-related diseases like angina, in which pain is caused by a reduction in blood supply to the heart muscle.

Chest Discomfort: Various sensations, including pressure, burning, or disquiet in the chest, are included in chest discomfort. It may be a sign of underlying health problems, such as heart difficulties, calling for prompt medical attention.

Pain in the Neck, Throat, Jaw, Upper belly area, or back

Challenge #1: Limited Annotated Data

Time-consuming and requiring specialized knowledge is the annotation of medical images for cardiac disorders. It is frequently difficult to get enough annotated data to train deep learning models, which makes it difficult to create reliable algorithms.

Challenge #2: Improve the Accurate Results of the Patient Variability:

Heart disease symptoms might differ considerably between patients and even over time within the same patient. To provide precise and trustworthy predictions, machine learning algorithms must consider these variations.

Challenge #3: Data Quality & Quantity

To develop precise models, it is essential to acquire large and varied medical picture datasets. Due to privacy issues, the wide range of imaging modalities, and the scarcity of labeled data, it is still difficult to acquire big, well-annotated datasets for cardiac disease prediction.

Challenge #4: Class Imbalance

Often, healthy cases exceed diseased cases in heart disease databases, indicating a class imbalance. To avoid bias and ensure accurate disease prediction, it is crucial to address this imbalance while training models.

Challenge #5: Data Augmentation

It is essential to produce a variety of data to enhance model generalization. However, it can be difficult to produce realistic and clinically useful synthetic images, especially for images used in medicine, where slight variations in the data can result in various diagnoses.

Challenge #6: Regulatory Daily Performance of Machine and Clinical Integration

Regulatory approval, validation, and establishment of their clinical utility are necessary for integrating machine learning and deep learning models into clinical practice. Models must adhere to medical norms and criteria to be adopted successfully.

Challenge #7: Real-Time Prediction of Heart Disease from Medical Images

For clinical decision-making, real-time prediction of heart disease from medical pictures is crucial. The technical difficulty of achieving low-latency predictions with high accuracy calls for effective model designs and hardware acceleration.

Challenge #8: Multi-Modality Integration

Various modalities, including MRI, CT, and echocardiography, are included in medical imaging. Due to variations in picture collection, resolution, and noise characteristics, combining data from many modalities to improve diagnostic accuracy is difficult.

Challenge #9: Complex Anatomical Structures

The heart has extensive anatomical features, making it a complicated organ. Technically, it is still difficult to create models that effectively classify and assess various features, including chambers, valves, and vessels.

Challenge #10: Data Privacy and Security

Patient data is sensitively contained in medical photographs. It is difficult to develop techniques that allow for efficient analysis while maintaining patient confidentiality and data security.

The main goal of the entire research project was to conduct an analysis and fully understand the easily accessible information that previous researchers had provided. By employing innovative qualitative research methods from a range of earlier research publications on Google Scholar, we can use the research project to investigate the problems with malware identification. Data collection and analysis are the core studies that are part of the research process. Finding high-quality data, however, produces incredible study findings.

The main objective of the ongoing research was to examine and address issues about data processing techniques. The purpose of the focused literature review was to investigate qualitative research study methodologies, the field of malware detection, and sparse empirical diagnosis. The associated community's cultural experiences, social experiences, and behavioral patterns can all be made clear and understood by applying the selected approach. This research study aims to investigate in-depth the impending issues related to plant diseases.

To carefully evaluate previous study findings and select the most appropriate data collection techniques based on the available facts, every research methodology necessitates data collection. Numerous techniques are used to collect data, such as forms, transactional tracking, Wikipedia, surveys, phone interviews, Kaggle, Google Scholar, and Google. However, determining the nature of unexplored land is a challenging task.

The process of data analysis A crucial step in the transcribing and content analysis process is approaching phrases or text. The qualitative content analysis, which usually employed a single theme to communicate the entire document, contained sentences, single words, and paragraphs. Apart from assessing theoretical hypotheses, the process of data analysis facilitates comprehension and advancement of data collection procedures.

The heart disease dataset, which consists of heart images used for detection, is accessible online through the Kaggle repository. Furthermore, training accounted for 70% of the data, with the remaining 30% being used for testing. Heart disease datasets that we have chosen contain a variety of properties, such as shape, texture, smoothness, and more.

4. **Results**

The classification algorithms track the results through a confusion matrix to improve and achieve the highest accuracy. With the help of a confusion matrix, we can easily find out the target accuracy.

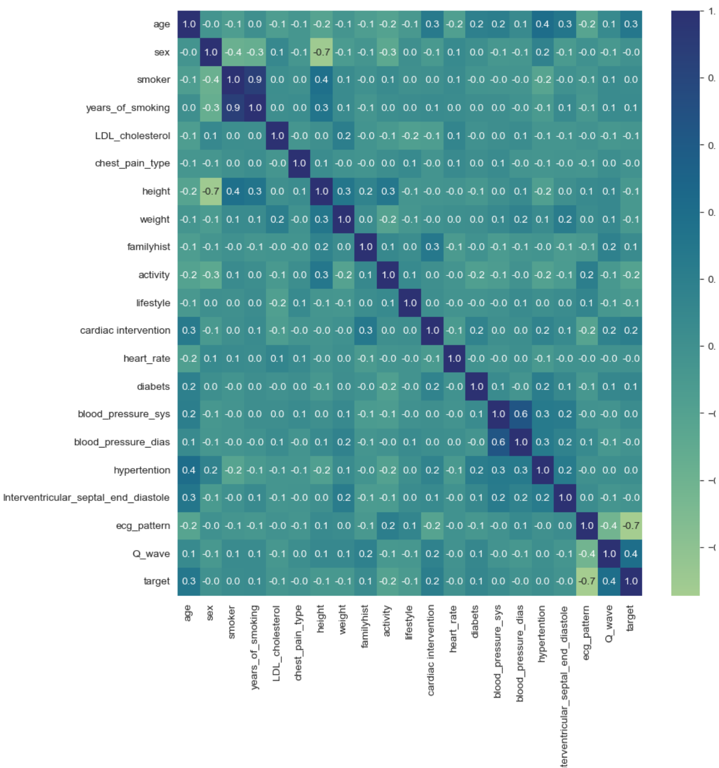

Figure # 2 Confusion Matrix of the Algorithms

In figure 2 confusion matrix, which shows a classification model's accuracy, is a machine learning performance evaluation tool. The quantity of false positives, false negatives, true positives, and true negatives is shown. This matrix helps with misclassification detection, predictive accuracy improvement, and model performance analysis. An N x N matrix, where N is the total number of target classes, is called a confusion matrix and is used to assess how well a classification model performs. The machine learning model's projected values are compared with the actual target values in the matrix. This provides us with a comprehensive understanding of the types of mistakes and performance metrics of our classification model. if there are more than two classes in your data. You may obtain an 80% classification accuracy with three or more classes, but you won't know if this is because the model is predicting each class equally well or if it is ignoring one or two classes.

When the number of classes in your data is not even. If 90 out of every 100 records belong to a single class, even though you may still get an accuracy of 90% or more, this is not a good result. You can improve your score by consistently guessing the most common class value.

The sum of the numbers in the men column (3 + 2) is the total number of genuine males in the dataset.

The total of the numbers in the women column (1 + 4) represents the number of genuine women in the dataset.

The correct values in the matrix (3 + 4) are arranged in a diagonal line from top left to bottom right.

Predicting men as women resulted in more mistakes than predicting women as men.

In Figure 3 Feature extraction is the process of condensing a massive amount of data into a small number of relevant fragments. Feature selection is the process of deciding the subset of relevant attributes to utilize when creating statistical and machine-learning models. Reducing the amount of information speeds up the machine learning and generalization processes and helps the machine create the model more quickly. Descriptive statistics for the given dataset Table 48 has 64 values for age, 48 values for heart rate, 48 values for blood pressure (both systolic and diastolic), and 66 values for the target. A portion of the 66 patients' heart rate, bp_diastolic, and bp_systolic variables are missing. As a result, before building the model, this study employed several techniques to fill in the missing variables. The majority of the patients in this study are elderly, with a mean age of 59. The age ranges that applied were minimum (28 years) and maximum (92 years). There are no populations with low ages, as indicated by the data's left-skewed age distribution.

5. CNN Model Performance

Regarding CNN model accuracy and epoch, it's important to note that the accuracy of the model is influenced by the quality and representativeness of the dataset, the complexity of the problem, the architecture of the CNN model, and the optimization process. The choice of the appropriate number of epochs during training is crucial to prevent overfitting or underfitting of the model. It involves finding a balance between the model's ability to learn from the data and generalize to unseen examples. The significance of each attribute in terms of model accuracy and epoch can vary depending on the specific dataset and problem at hand. Some attributes, such as age, sex, smoking status, and blood pressure, may have a strong influence on model accuracy and require fewer epochs to learn meaningful patterns. On the other hand, attributes with less direct associations with cardiovascular disease, such as height and weight, may have a relatively smaller impact on model accuracy and may require more epochs to extract relevant features.

Model Accuracy: The y-axis represents the accuracy of the model, ranging from 0 to 1.0. The starting accuracy value is 0, indicating no correct predictions initially. The ending accuracy value is 1.0, representing perfect accuracy. The intervals between the accuracy values on the y-axis are 0.1, showing increments of 0.1.

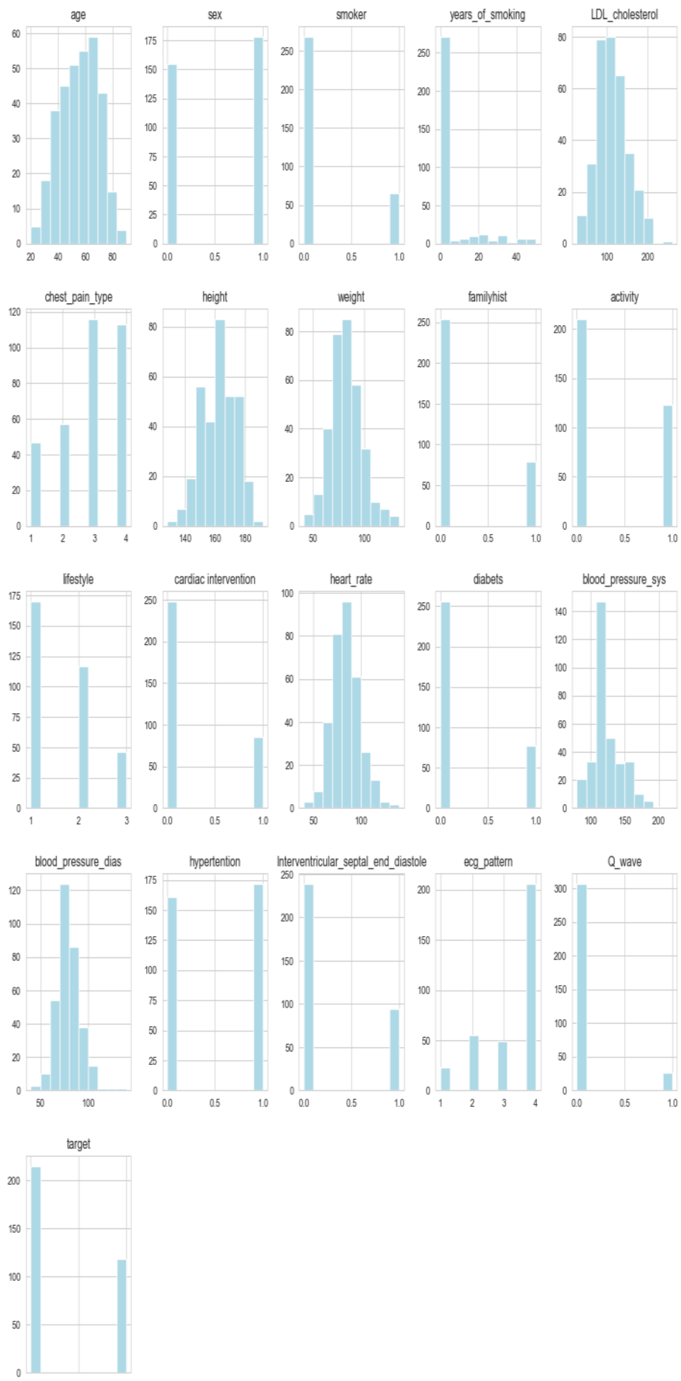

Figure # 3 Frequency Distribution of all Features

Model Loss: The y-axis represents the loss of the model, ranging from 0.7 to 0.1. The starting loss value is 0.7, indicating a high initial loss. The ending loss value is 0.1, representing a low loss value. The intervals between the loss values on the y-axis are 0.1, showing a decrease in loss by 0.1. Now we described some important attributes.

1. **Age**:
   - Accuracy: The graph will show the accuracy of the model in predicting age values. As the number of epochs increases, the model's accuracy in predicting age is expected to improve gradually. The accuracy may start from 0 and progressively increase, potentially reaching 1.0 or close to it.
   - Loss: The graph will show the loss incurred by the model when predicting age values. Initially, the loss may be relatively high, but it should decrease over epochs as the model learns from the data.
2. **Gender:**
   - Accuracy: The graph will show the accuracy of the model in predicting the sex attribute. The accuracy should increase over epochs, indicating the model's ability to predict sex correctly. The accuracy values may vary, but they should approach the maximum value of 1.0.
   - Loss: The graph will show the loss incurred by the model when predicting the sex attribute. The loss should decrease as the model learns the patterns in the data and improves its predictions.
3. **Smoke:**
   - Accuracy: The graph will show the accuracy of the model in predicting smoking habits. The accuracy should increase over epochs as the model learns to differentiate between smokers and non-smokers. The accuracy values may fluctuate but should generally show an upward trend.
   - Loss: The graph will show the loss incurred by the model when predicting smoking habits. The loss should decrease as the model becomes more proficient at identifying smoking patterns in the data.
4. **Years:**
   - Accuracy: The graph will show the accuracy of the model in predicting the number of years since diagnosis or follow-up. The accuracy should improve over epochs as the model learns to estimate the years accurately. The accuracy values may gradually increase and approach the maximum value of 1.0.
   - Loss: The graph will show the loss incurred by the model when predicting the number

of years since diagnosis or follow-up. Initially, the loss may be high, but it should decrease over epochs as the model adjusts its parameters to minimize the error.

5. **LDL (Low-Density Lipoprotein):**
   - Accuracy: The graph will show the accuracy of the model in predicting LDL cholesterol levels. The accuracy should improve as the model learns to predict LDL levels more accurately. The accuracy values may increase steadily or fluctuate, but they should generally show an upward trend.
   - Loss: The graph will show the loss incurred by the model when predicting LDL cholesterol levels. The loss should decrease over epochs as the model adjusts its weights to minimize the error between predicted and actual LDL levels.

The specific values of accuracy and loss for each attribute and epoch cannot be provided without the actual training and evaluation of the CNN model on the dataset. The graphs would vary depending on the complexity of the data, the quality of the model architecture, and the training process.

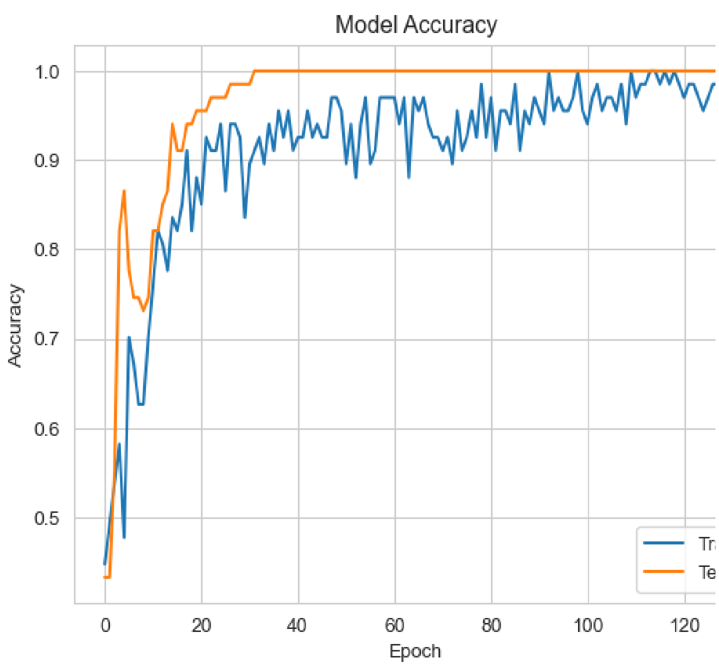

Figure # 4 Model Accuracy

The most basic and natural evaluation criteria for algorithms are accuracy and error in Figure 4. The percentage of accurate predictions or classifications produced by the algorithm is measured by accuracy, whilst the percentage of wrong ones is measured by error. These metrics are appropriate for binary or multiclass issues when the classes are balanced and the cost of misclassification is equal. They are also simple to compute and analyze. They may, however, be deceptive or insufficient in situations when there is an imbalance in the classes, the cost of incorrect classification is unpredictable, or the algorithm produces probabilities or scores rather than discrete labels.

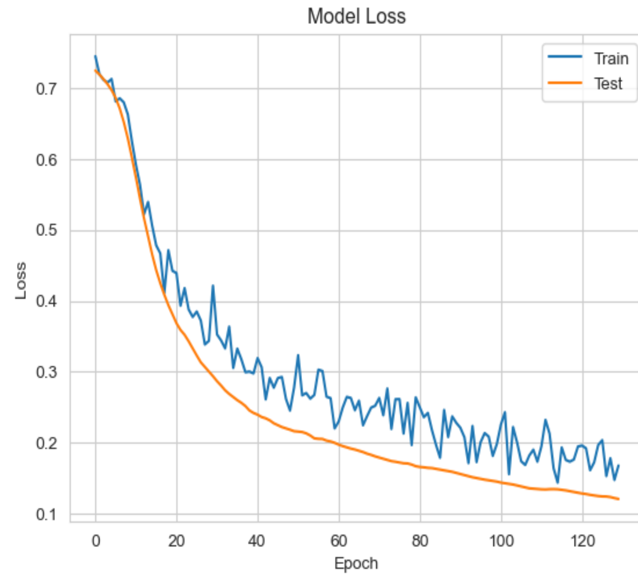

Figure # 5 Model Loss

Figure 5 represents the Deep learning model, the difference between expected and actual values is quantified by a mathematical function called a loss function. It gauges the model's effectiveness and directs the optimization procedure by indicating how well the model fits the data. However, "Cost function" and "loss function" are frequently used synonymously in deep learning. The notion of a function that determines the error or difference between expected and actual values is the same in each of them. To increase accuracy, the cost or loss function is reduced while the model is being trained.

4.1 CNN Model Performance

What actions can I take to enhance these Models' performance? More precisely, what are the few essential guidelines that one should constantly strive to adhere to to enhance a Deep Learning Model's efficiency? It's not incorrect. CNN needs to be able to automatically extract features from the data, which is typically only feasible in situations when a large amount of training data is available.

```
3/3 [==============================] - 0s 987us/step
Results for Categorical Model
1.0
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        38
           1       1.00      1.00      1.00        29

    accuracy                           1.00        67
   macro avg       1.00      1.00      1.00        67
weighted avg       1.00      1.00      1.00        67
```

Figure # 6 Performance CNN Models

The performance of the CNN model can be enhanced by adjusting parameters such as learning rate and epochs.

Performance is undoubtedly impacted by the number of epochs Figure 6. Performance improves after a significant number of epochs. However, some testing is required to determine the learning rate and epochs. It is evident that training accuracy does not increase and training loss does not decrease after a given number of epochs. We can choose the number of epochs as a result. The CNN model's dropout layer is another option. During model compilation, the appropriate optimizer must be chosen based on the application. We can employ other optimizers, such as SGD, rmsprop, etc. The model must be tuned using a variety of optimizers. Each of these factors has an impact on CNN's performance.

4.2 SVM Algorithm Performance

Support Vector Machine (SVM) is a supervised learning algorithm that is commonly used for solving classification problems. Figure #5.6 shows that the objective of SVM is to identify the variables with the closest margins with the help of a line and cleanly separate two classes of data points. There are many ways to solve the SVM problems, it's important to identify the gap among objects. The Support Vector Machine (SVM) technique is well known for completing a variety of classification tasks with outstanding results. Due to its proficiency in handling complicated datasets and capacity to identify the best decision limits, the accuracy of its results is particularly impressive.

```
confussion matrix
[[31  7]
 [ 2 27]]


Accuracy of Support Vector Classifier: 86.56716417910447

              precision    recall  f1-score   support

           0       0.94      0.82      0.87        38
           1       0.79      0.93      0.86        29

    accuracy                           0.87        67
   macro avg       0.87      0.87      0.87        67
weighted avg       0.88      0.87      0.87        67
```

Figure # 7 SVM Algorithm Performance

SVM aims to increase the margin between several classes, which results in strong generalization and decreased overfitting. SVM's accuracy in real-world applications is frequently credited with its ability to handle high-dimensional data, nonlinear correlations, and even situations with few training samples. SVM can identify subtle patterns that could be problematic for other algorithms by translating data into a higher-dimensional space. Furthermore, SVM's ability to use kernel functions makes accurate classification possible by transforming data into regions where connections are easier to see. The choice of suitable hyperparameters, such as the kernel and regularization parameters, has a significant impact on the accuracy of SVM that may be accomplished. SVM is a popular option in areas like image recognition, text classification, and bioinformatics because it consistently produces excellent accuracy rates across a variety of applications with the help of careful tweaking and cross-validation.

4.3 Decision Tree Algorithm Performance

The decision tree algorithm is well known in machine learning techniques for regression and classification problems the usage of the tree. The most useful classifiers are supervised learning including input data into subset recursively succeeding nodes to reach the conclusions. Each partition or decision is represented as a node in the tree-like structure. It is very important first of all focus on dataset data preparation and analysis associated with target values (numerical values or class labels for classification and regression to predict the outcomes.

```
confussion matrix
[[36  2]
 [ 0 29]]


Accuracy of DecisionTreeClassifier: 97.01492537313433

              precision    recall  f1-score   support

           0       1.00      0.95      0.97        38
           1       0.94      1.00      0.97        29

    accuracy                           0.97        67
   macro avg       0.97      0.97      0.97        67
weighted avg       0.97      0.97      0.97        67
```

Figure # 8 Decision Tree Algorithms Performance

Feature selection methods are necessary to select threshold values, the chosen feature divides the data into two or more parts. To create a prediction for a new data point, start the path from the root node to the leaf node to find out the feature values of the data point. The final prediction values are associated with the leaf node of the class label classification and regression.

4.4 KNN Algorithm Performance

KNN stands for k-nearest neighbors' algorithms and it is used for both classification and regression problems. KNN algorithms can classify the data and measure the similarity of new data points. However, KNN is more reliable as compared to other algorithms for finding accurate results of the data. Most of the researchers already implemented the KNN algorithm and gained accuracy through the feature selection problem. The implementation of KNN algorithms needs some important Python libraries which are easily gained the optimal solution.

```
confussion matrix
[[35  3]
 [13 16]]


Accuracy of K-NeighborsClassifier: 76.11940298507463

              precision    recall  f1-score   support

           0       0.73      0.92      0.81        38
           1       0.84      0.55      0.67        29

    accuracy                           0.76        67
   macro avg       0.79      0.74      0.74        67
weighted avg       0.78      0.76      0.75        67
```

Figure # 9 KNN Algorithms Performance

Suppose there are two categories, i.e., Category A and Category B, and we have a new data point x1, so this data point will lie in which of these categories? To solve this type of problem, we need a K-NN algorithm. With the help of K-NN, we can easily identify the category or class of a particular dataset.

4.5 XGB Classifier (Extreme Gradient Boost) Performance

Extreme Gradient Boost, or XGBoost, is a powerful and well-known machine learning method that excels at a variety of classification problems. Its outcomes stand out for their outstanding robustness, accuracy, and capacity for handling complicated information. The XGBoost classifier's boosting technique is one of the main elements in the accuracy of its excellent results. By integrating several weak learners (often decision trees), XGBoost uses an ensemble learning strategy to build a powerful predictive model. It continuously improves its predictions and lowers mistakes by focusing on situations that earlier models incorrectly identified. This iterative process of boosting improves accuracy and aids the algorithm's ability to recognize complex patterns in the input.

```
confussion matrix
[[32  6]
 [ 3 26]]


Accuracy of Extreme Gradient Boost: 86.56716417910447

              precision    recall  f1-score   support

           0       0.91      0.84      0.88        38
           1       0.81      0.90      0.85        29

    accuracy                           0.87        67
   macro avg       0.86      0.87      0.86        67
weighted avg       0.87      0.87      0.87        67
```

Figure # 10XGB Classifier (Extreme Gradient Boost) Performance

The accuracy of XGBoost is further improved by its capacity to handle both linear and non-linear correlations within the data. The technique is appropriate for datasets with complicated underlying structures because it can automatically find and capture complex connections among variables. Furthermore, the regularization approaches used in the model prevent overfitting and guarantee that the model generalizes well to new data. The optimization approach used by XGBoost is another element that contributes to its astounding accuracy. It makes use of a gradient boosting framework to reduce a loss function, which successfully directs the algorithm to the most advantageous model parameters. The model is updated iteratively by XGBoost to improve prediction accuracy depending on the gradient of the loss function.

4.6 Logistic Regression Performance

Logistic regression is a classification algorithm based on some significant dependent variables also used to predict certain classes' probability. The ability of the logistic regression model to collect the input features bias and in most cases measure the main results of algorithms. This algorithm concludes the outcome through the classification method and evaluating the dataset features including one or more dependent and independent variables. The outcome of the algorithm measure and extract two possibilities from the logistic regression.

```
confussion matrix
[[32  6]
 [ 2 27]]


Accuracy of Logistic Regression: 88.05970149253731

              precision    recall  f1-score   support

           0       0.94      0.84      0.89        38
           1       0.82      0.93      0.87        29

    accuracy                           0.88        67
   macro avg       0.88      0.89      0.88        67
weighted avg       0.89      0.88      0.88        67
```

Figure # 11 Logistic Regression Performance

4.7 Overall Performance Results

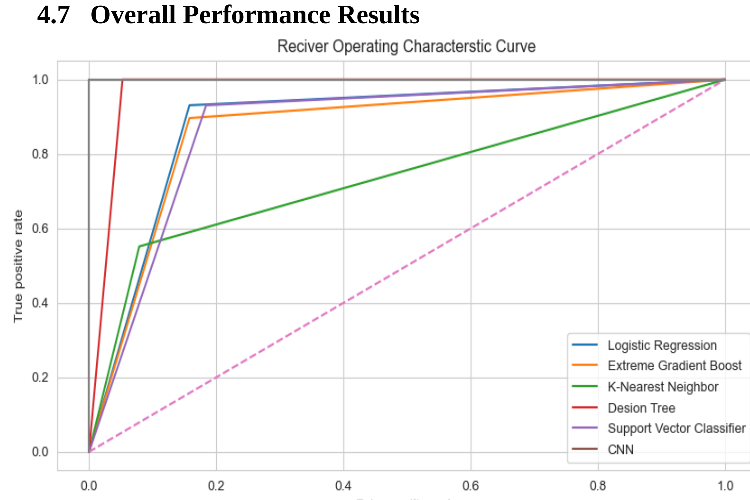

Figure # 12 Overall Performance Results

To compare the machine learning and deep learning algorithms with each other figure 12. The overall results are impressive but deep learning accuracy is high as compared to machine learning classifiers. The analysis of classical machine and deep learning techniques based on confusion matrix methods. The results of logistic regression, decision tree, CNN model, and SVM model achieved the target accuracy on heart disease datasets. These outcomes show how well the conventional neural network; at 88.069%, had the highest accuracy. After that, the Logistic Regression algorithm also yielded comparable results. Decision trees, on the other hand, performed the highest score, scoring 97.0149% accuracy, 76.1194% less than KNN Algorithms. According to the whole results decision tree highest accuracy rate.

4.8 Reduce Overfitting Problems in Feature Selection

Overfitting occurs when our algorithms fit training data too well, while it becomes difficult to analyze the training data through the models to generalize. The training data should recognize related patterns in MRI images. With more training data for algorithms, the number of problems generated increased. The algorithms were confused because there were many issues in your training data, like resized images, features, shapes, colors, and textures of data, etc. Overfitting problems are caused by statistical models learning from and training on the data to make noise. Empty and irreverent features are included in the dataset feature selections. Overfitting hurts the models' performance and also reduces the accuracy of new data. Data becomes noisy during model training, and irrelevant features of datasets pick up on the fluctuation. These kinds of issues arise in brain tumor image datasets when overfitting nonlinear and nonparametric functions to target the function constraint and the model limit. ML model to increase accuracy with the help of feature processing, feature engineering techniques, and model parameter tuning. However, time complexity is reduced by algorithms.

4.9 Improve Accuracy and Reduce Complexity of Algorithms

It is important to understand data clearly and also examine how each feature of data affects accuracy. Data generation and analysis are critical in any field of study. With the help of accurate data, we can easily reach our target. The first step in improving accuracy is to select high-quality data, analyze it to remove unnecessary and incorrect information, and then try to fix it. Before the experiment, set quality data goals based on the model's results. Data review is the best way to reduce errors, avoid overloading, and maintain data standards. So we know that algorithms are nothing more than rules and regulations for problem-solving. Additionally, reduces the algorithm's time complexity while focusing on its space complexity. Complexity mostly depends on the better logic of any problems and efficient models' performance. If you want to reduce the complexity of algorithms, you should first focus on what the problem is, and then create logic for optimal algorithm results. There are various methods for reducing the complexity of classifiers, but researchers must determine which method is best for our problem.

5. Conclusions and Future Work

Machine learning techniques, such as data mining tools, have been successfully applied to improve the performance of medical diagnosis for heart disease. Heart disease is a major global health concern and a leading cause of death worldwide. It encompasses various forms, including coronary artery disease, heart failure, arrhythmias, valvular heart disorders, and congenital heart defects. Early detection and risk assessment are crucial in managing heart disease. Advanced imaging techniques and diagnostic technologies play a vital role in identifying heart disease at an early stage, enabling timely interventions and improved outcomes. The study emphasized the need for a comprehensive understanding of heart disease to develop focused and efficient interventions that can reduce its global impact and improve cardiovascular health for everyone using machine learning and deep learning algorithms. Furthermore, this research can explore the use of machine learning techniques and data mining tools to enhance the classification and prediction of heart disease, improving the accuracy and efficiency of medical diagnosis. Future studies can focus on investigating the underlying causes and risk factors of heart disease in more detail, including the interaction between genetic predisposition and environmental influences. This can help in developing targeted interventions and preventive measures. There is a need for ongoing research to develop and refine advanced imaging techniques and diagnostic technologies for early detection and risk assessment of heart disease. This can enable healthcare professionals to take timely action and prevent irreversible damage. Further exploration of non-invasive methods for efficient healthcare system diagnosis can be a valuable area of future research. This can involve the use of machine learning as well as deep learning techniques to classify and analyze large amounts of data generated in hospitals and medical institutes.

References

[1] Tougui, Ilias, Abdelilah Jilbab, and Jamal El Mhamdi. "Heart disease classification using data mining tools and machine learning techniques." *Health and Technology* 10 (2020): 1137-1144.

[2] Muppalaneni, Naresh Babu, Maode Ma, Sasikumar Gurumoorthy, R. Kannan, and V. Vasanthi. "Machine learning algorithms with ROC curve for predicting and diagnosing the heart disease." *Soft computing and medical bioinformatics* (2019): 63-72.

[3] Gonsalves, Amanda H., Fadi Thabtah, Rami Mustafa A. Mohammad, and Gurpreet Singh. "Prediction of coronary heart disease using machine learning: an experimental analysis." In *Proceedings of the 2019 3rd International Conference on Deep Learning Technologies*, pp. 51-56. 2019.

[4] Haq, Amin Ul, Jian Ping Li, Muhammad Hammad Memon, Shah Nazir, and Ruinan Sun. "A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms." *Mobile Information Systems* 2018 (2018): 1-21.

[5] Bharti, Rohit, Aditya Khamparia, Mohammad Shabaz, Gaurav Dhiman, Sagar Pande, and Parneet Singh. "Prediction of heart disease using a

combination of machine learning and deep learning." *Computational intelligence and neuroscience* 2021 (2021).

[6] Mohan, Senthilkumar, Chandrasegar Thirumalai, and Gautam Srivastava. "Effective heart disease prediction using hybrid machine learning techniques." *IEEE Access* 7 (2019): 81542-81554.

[7] Kavitha, M., G. Gnaneswar, R. Dinesh, Y. Rohith Sai, and R. Sai Suraj. "Heart disease prediction using hybrid machine learning model." In *2021 6th international conference on Inventive Computation Technologies (ICICT)*, pp. 1329-1333. IEEE, 2021.

[8] Anjaneyulu, M., Divya Priya Degala, P. Devika, and V. Hema. "Effective heart disease prediction using hybrid machine learning techniques." In *AIP Conference Proceedings*, vol. 2492, no. 1. AIP Publishing, 2023.

[9] Limbitote, Mangesh, Kedar Damkondwar, Dnyaneshwari Mahajan, and Pushkar Patil. "A survey on prediction techniques of heart disease using machine learning." *International Journal of Engineering Research & Technology (IJERT)* 9, no. 6 (2020): 2278-0181.

[10] Sharanyaa, S., S. Lavanya, M. R. Chandhini, R. Bharathi, and K. Madhulekha. "Hybrid Machine Learning Techniques for Heart Disease Prediction." *International Journal of Advanced Engineering Research and Science* 7, no. 3 (2020).

[11] Chitra, R., and V. Seenivasagam. "Review of heart disease prediction system using data mining and hybrid intelligent techniques." *ICTACT journal on soft computing* 3, no. 04 (2013): 605-09.

[12] Bui, Anh L., Tamara B. Horwich, and Gregg C. Fonarow. "Epidemiology and risk profile of heart failure." *Nature Reviews Cardiology* 8, no. 1 (2011): 30-41.

[13] Heidenreich, Paul A., Justin G. Trogdon, Olga A. Khavjou, Javed Butler, Kathleen Dracup, Michael D. Ezekowitz, Eric Andrew Finkelstein, et al. "Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association." *Circulation* 123, no. 8 (2011): 933-944.

[14] Dangare, Chaitrali S., and Sulabha S. Apte. "Improved study of heart disease prediction system using data mining classification techniques." *International Journal of Computer Applications* 47, no. 10 (2012): 44-48.

[15] Kavitha, R., and E. Kannan. "An efficient framework for heart disease classification using feature extraction and feature selection technique in data mining." In *2016 international conference on emerging trends in engineering, technology, and science (octets)*, pp. 1-5. IEEE, 2016.

[16] Pahwa, Kanika, and Ravinder Kumar. "Prediction of heart disease using hybrid technique for selecting features." In *2017 4th IEEE Uttar Pradesh section international conference on electrical, computer and electronics (UPCON)*, pp. 500-504. IEEE, 2017.

[17] Bouali, Hanen, and Jalel Akaichi. "Comparative study of different classification techniques: heart disease use case." In *2014 13th International Conference on Machine Learning and Applications*, pp. 482-486. IEEE, 2014.

[18] Pouriyeh, Seyedamin, Sara Vahid, Giovanna Sannino, Giuseppe De Pietro, Hamid Arabnia, and Juan Gutierrez. "A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease." In *2017 IEEE symposium on computers and communications (ISCC)*, pp. 204-207. IEEE, 2017.

[19] Pugazhenthi, D., and V. Meenakshi. "Detection of ischemic heart diseases from medical images." In *2016 International Conference on Micro-Electronics and Telecommunication Engineering (ICMETE)*, pp. 355-360. IEEE, 2016.

[20] Saini, Ridhi, Namita Bindal, and Puneet Bansal. "Classification of heart diseases from ECG signals using wavelet transform and kNN classifier." In *International Conference on Computing, Communication & Automation*, pp. 1208-1215. IEEE, 2015.

[21] Ekız, Simge, and Pakize Erdoğmuş. "Comparative study of heart disease classification." In *2017 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*, pp. 1-4. IEEE, 2017.

[22] Chauhan, Renu, Pinki Bajaj, Kavita Choudhary, and Yogita Gigras. "Framework to predict health diseases using attribute selection mechanism." In *2015 2nd international conference on computing for sustainable global development (INDIACom)*, pp. 1880-1884. IEEE, 2015.

[23] Jabbar, M. A., B. L. Deekshatulu, and Priti Chndra. "Alternating decision trees for early diagnosis of heart disease." In *International conference on circuits, communication, control and computing*, pp. 322-328. IEEE, 2014.

[24] Moradi, Hamed, Akram Al-Hourani, Gianmarco Concilia, Farnaz Khoshmanesh, Farhad R. Nezami, Scott Needham, Sara Baratchi, and Khashayar Khoshmanesh. "Recent developments in modeling, imaging, and monitoring of cardiovascular diseases using machine learning." *Biophysical Reviews* 15, no. 1 (2023): 19-33.

[25] Divya, K., Akash Sirohi, Sagar Pande, and Rahul Malik. "An IoMT assisted heart disease diagnostic system using machine learning techniques." *Cognitive internet of medical things for smart healthcare: services and applications* (2021): 145-161.

[26] Ramprakash, P., R. Sarumathi, R. Mowriya, and S. Nithyavishnupriya. "Heart disease prediction using deep neural network." In *2020 International Conference on Inventive Computation Technologies (ICICT)*, pp. 666-670. IEEE, 2020.

[27] Gárate-Escamila, Anna Karen, Amir Hajjam El Hassani, and Emmanuel Andrès. "Classification models for heart disease prediction using feature selection and PCA." *Informatics in Medicine Unlocked* 19 (2020): 100330.

[28] Santhanam, T., and E. P. Ephzibah. "Heart disease classification using PCA and feed forward neural networks." In Mining Intelligence and Knowledge Exploration: First International Conference, MIKE 2013, Tamil Nadu, India, December 18-20, 2013. Proceedings, pp. 90-99. Springer International Publishing, 2013.

[29] Amma, NG Bhuvaneswari. "Cardiovascular disease prediction system using genetic algorithm and neural network." In *2012 international conference on computing, communication and applications*, pp. 1-5. IEEE, 2012.

[30] Srinivas, K., G. Raghavendra Rao, and A. Govardhan. "Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques." In *2010 5th International Conference on Computer Science & Education*, pp. 1344-1349. IEEE, 2010.

[31] Das, Resul, Ibrahim Turkoglu, and Abdulkadir Sengur. "Effective diagnosis of heart disease through neural networks ensembles." *Expert systems with applications* 36, no. 4 (2009): 7675-7680.

[32] Sen, Sanjay Kumar. "Predicting and diagnosing of heart disease using machine learning algorithms." *International Journal of Engineering and Computer Science* 6, no. 6 (2017): 21623-21631.

[33] Khan, Shabia Shabir, and S. M. K. Quadri. "Prediction of angiographic disease status using rule based data mining techniques." In *Biol. Forum Int. J*, vol. 8, no. 2, pp. 103-107. 2016.

[34] Ratnasari, N. R., Adhi Susanto, and Indah Soesanti. "Thoracic X-ray features extraction using thresholding-based ROI template and PCA-based features selection for lung TB classification purposes." In *2013 3rd International Conference on Instrumentation, Communications, Information Technology and Biomedical Engineering (ICICI-BME)*, pp. 65-69. IEEE, 2013.

[35] Susanto, Adhi, and Indah Soesanti. "Thoracic X-ray features extraction using thresholding-based ROI template and PCA-based features selection for lung TB classification purposes." In *2013 3rd International Conference on Instrumentation, Communications, Information Technology and Biomedical Engineering (ICICI-BME)*.

[36] Kamencay, Patrik, Robert Hudec, Miroslav Benco, and Martina Zachariasova. "Feature extraction for object recognition using PCA-KNN with application to medical image analysis." In *2013 36th International Conference on Telecommunications and Signal Processing (TSP)*, pp. 830-834. IEEE, 2013.

[37] Avendano-Valencia, D., F. Martinez-Tabares, D. Acosta-Medina, I. Godino-Llorente, and G. Castellanos-Dominguez. "TFR-based feature extraction using PCA approaches for discrimination of heart murmurs." In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5665-5668. IEEE, 2009.

[38] Zhang, Dejun, Lu Zou, Xionghui Zhou, and Fazhi He. "Integrating feature selection and feature extraction methods with deep learning to predict clinical outcome of breast cancer." *Ieee Access* 6 (2018): 28936-28944.

[39] Rajagopal, Rekha, and Vidhyapriya Ranganathan. "Evaluation of effect of unsupervised dimensionality reduction techniques on automated arrhythmia classification." *Biomedical Signal Processing and Control* 34 (2017): 1-8.

[40] Yaghouby, F., A. Ayatollahi, and R. Soleimani. "Classification of cardiac abnormalities using reduced features of heart rate variability signal." *World Applied Sciences Journal* 6, no. 11 (2009): 1547-1554.

[41] Singh, Ram Sewak, Barjinder Singh Saini, and Ramesh Kumar Sunkaria. "Detection of coronary artery disease by reduced features and extreme learning machine." *Clujul Medical* 91, no. 2 (2018): 166.

[42] Dun, Boyang, Eric Wang, and Sagnik Majumder. "Heart disease diagnosis on medical data using ensemble learning." *Comput. Sci* 1 (2016): 1-5.

[43] Ani, R., Aneesh Augustine, N. C. Akhil, and O. S. Deepa. "Random forest ensemble classifier to predict the coronary heart disease using risk factors." In *Proceedings of the International Conference on Soft Computing Systems: ICSCS 2015, Volume 1*, pp. 701-710. Springer India, 2016.

[44] Krishnani, Divya, Anjali Kumari, Akash Dewangan, Aditya Singh, and Nenavath Srinivas Naik. "Prediction of coronary heart disease using supervised machine learning algorithms." In *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*, pp. 367-372. IEEE, 2019.

[45] Verma, Luxmi, Sangeet Srivastava, and P. C. Negi. "A hybrid data mining model to predict coronary artery disease cases using non-invasive clinical data." *Journal of medical systems* 40 (2016): 1-7.

[46] Jabbar, M. A., B. L. Deekshatulu, and Priti Chandra. "Intelligent heart disease prediction system using random forest and evolutionary approach." *Journal of network and innovative computing* 4, no. 2016 (2016): 175-184.

[47] Shorewala, Vardhan. "Early detection of coronary heart disease using ensemble techniques." *Informatics in Medicine Unlocked* 26 (2021): 100655.

[48] KL, Santhosh Kumar, and J. A. Madhurya. "Coronary Artery Disease Prediction using Data Mining Techniques." In *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, pp. 693-697. IEEE, 2020.

[49] Pal, Madhumita, and Smita Parija. "Prediction of heart diseases using random forest." In *Journal of Physics: Conference Series*, vol. 1817, no. 1, p. 012009. IOP Publishing, 2021.

[50] Kumar, M. Nikhil, K. V. S. Koushik, and K. Deepak. "Prediction of heart diseases using data mining and machine learning algorithms and tools." *International Journal of Scientific Research in Computer Science, Engineering and Information Technology* 3, no. 3 (2018): 887-898.

[51] Hassan, Ch Anwar ul, Jawaid Iqbal, Rizwana Irfan, Saddam Hussain, Abeer D. Algarni, Syed Sabir Hussain Bukhari, Nazik Alturki, and Syed Sajid Ullah. "Effectively predicting the presence of coronary heart disease using machine learning classifiers." *Sensors* 22, no. 19 (2022): 7227.

[52] Jackins, V., S. Vimal, Madasamy Kaliappan, and Mi Young Lee. "AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes." *The Journal of Supercomputing* 77 (2021): 5198-5219.

[53] Tasnim, Farzana, and Sultana Umme Habiba. "A comparative study on heart disease prediction using data mining techniques and feature selection." In *2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, pp. 338-341. IEEE, 2021.

[54] Muhammad, L. J., Ahmed Abba Haruna, Ibrahim Alh Mohammed, Mansir Abubakar, Bature Garba Badamasi, and Jamila Musa Amshi. "Performance evaluation of classification data mining algorithms on coronary artery disease dataset." In *2019 9th International Conference on Computer and Knowledge Engineering (ICCKE)*, pp. 1-5. IEEE, 2019.

[55] Abdullah, A. Sheik, and R. Rajalaxmi. "A data mining model for predicting the coronary heart disease using random forest classifier." In *International Conference in Recent Trends in Computational Methods, Communication and Controls*, pp. 22-25. 2012.

[56] Padmaja, B., Chintala Srinidhi, Kotha Sindhu, Kalali Vanaja, N. M. Deepika, and E. Krishna Rao Patro. "Early and accurate prediction of heart disease using machine learning model." *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 12, no. 6 (2021): 4516-4528.

[57] Elsayed, Hoda Ahmed Galal, and Liyakathunisa Syed. "An automatic early risk classification of hard coronary heart diseases using framingham scoring model." In *Proceedings of the Second International Conference on Internet of Things, Data and Cloud Computing*, pp. 1-8. 2017.

[58] Limbitote, Mangesh, Kedar Damkondwar, Dnyaneshwari Mahajan, and Pushkar Patil. "A survey on prediction techniques of heart disease using machine learning." *International Journal of Engineering Research & Technology (IJERT)* 9, no. 6 (2020): 2278-0181.

[59] Kavitha, M., G. Gnaneswar, R. Dinesh, Y. Rohith Sai, and R. Sai Suraj. "Heart disease prediction using hybrid machine learning model." In *2021 6th international conference on inventive computation technologies (ICICT)*, pp. 1329-1333. IEEE, 2021.

[60] Velusamy, Durgadevi, and Karthikeyan Ramasamy. "Ensemble of heterogeneous classifiers for diagnosis and prediction of coronary artery disease with reduced feature subset." *Computer Methods and Programs in Biomedicine* 198 (2021): 105770.

AUTHORS

First Author – Shah Hussain Bangash, MSSE, Iqra National University Peshawar, Pakistan,.

Second Author – Irfan Ullah Khan, MSSE, Iqra National University Peshawar, Pakistan,

Third Author – Zabi Ullah Khan, MSCS, Cecos University

Fourth Author Junaid Khan, MSCS, Iqra National University Peshawar, Pakistan,

Fifth Author Mohsin Tahir, Ph.D. Electrical**,** Iqra National University Peshawar, Pakistan

Sixth Author Waqas Ahmad, MSCS, Iqra National University Peshawar, Pakistan

Seventh Author -Ihtisham Ul Haq is Currently pursuing his Doctoral Degree in ICT and Computer Engineering from the University of Calabria, Italy**.**

8th Author: Maimoona Asad is currently at Shenzhen University's College of Electronics and Information Engineering, Maimoona Asad is pursuing a Ph.D. in Information and Communication Engineering.

9th Author Mohammad Mansoor Qadir.CECOS University of IT and Emerging Sciences, Peshawar.

Correspondence Author – Shah Hussain Bangash, MSSE, Iqra National University Peshawar, Pakistan (if any)