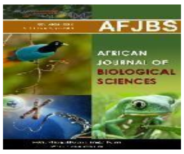
https://doi.org/10.33472/AFJBS.6.5.2024. 7576-7586



AfricanJournalofBiological Sciences

Machine Learning Techniques for Heart Disease Prediction: A Review

[1]Sandeep Kumar Mathariya, [2]Mahaveer jain, [3]Anil Patidar , [4]Shruti khete, [5]Payal Nimanpure, [6]Ayush Supath

[1,2,3]Assistant Professor, [4,5,6]Student

[1,2,3,4,5,6]Department of Computer Science & Engineering, Medic-Caps University Indore, MP , India

[1]mathariya@gmail.com,[2]profmahavir@gmail.com,[3]anilpatidar07@gmail.com,[4]shrutikhete1207@gmail.com, [5]payalnimanpure04@gmail.com,[6]ayushsupath13326@gmail.com,

Article History
Volume 6, Issue 5, 2024
Received: 15 May 2024
Accepted: 22 May 2024
doi:*10.33472/AFJBS.6.5.2024. 7576-7586*

Abstract—Over the past few years, heart disease has become one of the leading causes of death worldwide. Lifestyle, nutrition, work culture, etc. are changing all over the world, including developing countries, not developed countries. changes have contributed to this problem. Early detection of heart disease symptoms and ongoing medical care can help reduce the number of patients and ultimately reduce mortality. However, due to the limited number of medical facilities and specialists, it is difficult to provide regular patient care and counselling. Technology is needed to support patient care and treatment. Clinical data from multiple clinical procedures and continuous patient care can be used to develop predictive models for cardiovascular disease. Early diagnosis of heart disease can help inform lifestyle changes to reduce complications in high-risk patients, which may be important in medicine. This article reviews some of the most commonly used machine learning techniques to predict heart disease using clinical and historical data. Various techniques are discussed and compared. This report compares five methods published in the literature for predicting the timing of a heart attack.

Keywords— Neural network, machine learning, supervised learning, deep learning, healthcare, risk prediction.

1 Introduction

The heart is an important part of the same human body. And is responsible for the circulation of blood and supply of oxygen and nutrients to various tissues and organs. Any malfunction or disease affecting the heart can have serious consequence and can be fatal if not treated quickly. Factors such as lifestyle choices, stress and nutrition have a significant impact on heart disease. These diseases have become the leading cause of death worldwide, including in India. The burden of heart diseases not only affects people's health but also has health consequences such as increased healthcare expenses and decreased productivity. Accurate prediction and early detection of heart disease is important for effective management and prevention. Advances in medical technology and data analysis have led to development of predictive models that can assess individual risk factors and provide timely intervention. Addressing heart diseases requires a multifaceted approach, including public health measures to improve health, access to quality healthcare, and research is ongoing to better understand and treat heart disease. By emphasising prevention and early intervention, we can reduce the impact of heart disease and improve public health.

Over the years, heart disease has become a major concern. One of the most important problems in coronary artery disease is knowing the symptoms and treating the disease correctly. Early surgery is not sufficiently effective in predicting heart disease[1]. Many different clinical tools are available to predict coronary artery disease. These devices have two main problems; First, they are very expensive, and second, they are not very useful in calculating heart disease in humans. According to a recent review of WHO guidelines, medical professionals are only prepared to accurately predict 67% of coronary artery disease [2]. Therefore, there are many research methods to predict heart disease in humans. There are many different types of heart disease, and each has its own unique symptoms and treatment. For some people, lifestyle changes and medications can have a big impact on personal growth. For others, medical procedures may be needed for this to work [3].

It has a huge impact on the world population due to heart disease. GBD (Global Burden of Disease) 2019 is an international collaborative study designed to measure the burden of disease in all countries of the world. This study is an ongoing study, updated annually, and is designed to determine the relationship between age and gender and across regions from 1990 to 2019. This analysis results in better prediction of disease incidence, incidence, and mortality as well as health outcomes (e.g., disability-adjusted life years). Disability Adjusted Life Years (DALYs) are associated with impairment in activities of daily living and years of long-term loss and disability; life language can be assessed by normal assessment and disability; It may be in the form of a check or interest report[4]

2 Methodology

- **Data collection and progress:**

Blood sugar, heart rate, age, gender, cholesterol level, blood sugar etc. Collect etailed information with useful results such as. Care is taken to ensure that the data set is free of missing values, anomalies and inconsistencies. Standardize or institutionalize digital features for comparison.

- **Analytical Analysis :**

Perform statistical analysis to obtain the effects of events and discern patterns or relationships. Use data visualizations such as histograms, box plots, and line charts to gain insight into the relationship between variables.

- **Feature Selection:**

Using decision-making techniques such as correlation analysis, data extraction, or forward/backward selection to identify the most important factors in predicting heart disease. Select the behaviour that has the greatest impact on the target variable (proximity or absence of cardiovascular disease).

- **Model selection and training:**

Selecting data mining appropriate to the classification task, such as Decision Tree, Credulous Bayesian, Support Vector Machine (SVM) or Logistic regression. Split the dataset into a plan and test to evaluate and demonstrate performance. Use prepared data to train different models and tune their hyperparameters for better performance.

- **Model Evaluation:**

Evaluate the proposed model using metrics such as Accuracy, Accuracy, Testing, F1 Score and ROC-AUC to evaluate its performance in preventing heart diseases. Methods such as bracing can be used to increase the strength and stability of the structure.

- **Information Collection:**

Accumulate a comprehensive dataset containing data on individuals' statistic characteristics (age, sex), physiological estimations (blood weight, cholesterol levels), way of life variables (smoking propensities, work out recurrence), therapeutic history (diabetes, hypertension), and family history of heart malady. Datasets can be gotten from open stores, healthcare education, or collected through surveys.

- **Information Preprocessing:**

Handle lost values: Ascribe lost values utilizing procedures like cruel, middle, mode ascription, or progressed strategies like k-nearest neighbors (KNN) ascription or prescient modelling.

Exception discovery and treatment: Recognize exceptions utilizing measurable strategies (e.g., z-score, interquartile extent) and choose whether to evacuate them or apply change techniques.

Information normalization/standardization: Scale numerical highlights to a common extent to avoid certain highlights from ruling others amid demonstrated training.

Encoding categorical factors: Change over categorical factors into numerical representations utilizing strategies like one-hot encoding or name encoding.

- **Highlight Selection/Engineering:**

Perform exploratory information examination (EDA) to get the connections between highlights and the target variable (heart disease). Select instructive highlights based on space information, relationship examination, or include significance positioning calculations (e.g Arbitrary Woodland include importance). Build modern highlights that might capture imperative connections or intuitive between factors (e.g., BMI from stature and weight).

- **Show Selection:**

Select appropriate machine learning calculations for classification, considering components such as interpretability, adaptability, and performance. commonly utilized calculations for heart illness forecasts incorporate calculated relapse, choice trees, arbitrary woodlands, back vector machines (SVM), k-nearest neighbors (KNN), and neural networks.

- **Show Training:**

Part the dataset into preparing and testing sets (e.g., 70-30 or 80-20 proportion) to prepare and assess the models, respectively. Prepare the chosen models utilizing the preparing information, altering show parameters to minimize the chosen misfortune work (e.g., cross-entropy misfortune for calculated regression).

- **mi-Show Evaluation:**

Assess the prepared models utilizing suitable assessment measurements such as exactness, exactness, review, F1-score, and region beneath the ROC bend (AUC-ROC) on the test set. Survey the model's execution comprehensively to get its qualities and limitations.

- **Hyperparameter Tuning:**

Fine-tune the hyperparameters of the models utilizing strategies like lattice look, irregular look, or Bayesian optimization to optimize execution further. Hyperparameters incorporate parameters that are not learned amid preparation but influence the learning preparation, such as regularization quality, learning rate, or tree depth.

- **Demonstrate Validation:**

Approve the last to demonstrate utilizing strategies like k-fold cross-validation to guarantee its vigour and generalizability to inconspicuous data. This step makes a difference in evaluating how well the show performs on distinctive subsets of the information and gives more dependable gauges of performance.

- **Deployment:**

Send the prepared demonstration in a real-world setting, such as a healthcare framework, to anticipate heart malady hazard for unused individuals.Coordinated the show into the existing program foundation, guaranteeing compatibility and scalability.

- **Checking and Maintenance:**

Ceaselessly screen the model's execution in the generation environment to distinguish any debasement or drift. Overhaul the demonstration intermittently with unused information or progressed calculations to keep up its precision and significance over time. Conduct standard reviews to guarantee compliance with moral and administrative rules, counting information protection and security requirements.

- **Feature selection:**

Feature selection is a crucial step in the machine learning pipeline, especially when dealing with high-dimensional datasets like those often encountered in healthcare analytics, such as cardiac disease diagnosis. In this organization, tests were conducted with and without highlight choice to evaluate the impact of include determination.

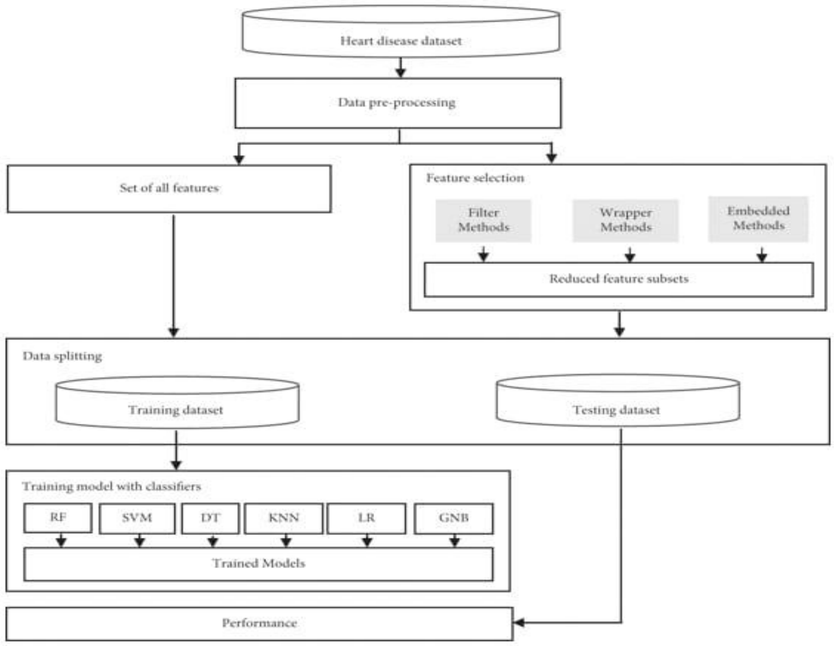


Figure:1 the Process representation

They include choice points to recognize the most critical highlights of cardiac infections. Besides, highlight choice makes a difference to develop a more exact show by dispensing with or underrepresenting the less important highlights, minimize preparing time and improving learning execution [10]. The behavior of a few include determination approaches beneath the three major categories (channel, wrapper, and implanted) is evaluated in this try, highlight choice methods having a place in three categories were independently connected to introductory datasets. Include determination methods start by making a subset, but that subset era depends on the sort of approach. outlines the handle of highlight choice taken after by each of the three categories to recognize ideal highlight subsets.

The channel strategy chooses the best subset promptly some time recently, passing it to the learning calculation. The remaining two approaches, wrapper and implanted, make the ideal subset in combination with the learning calculation. In addition to the other strategies, the implanted strategy consolidates the benefits of both the channel and wrapper methods

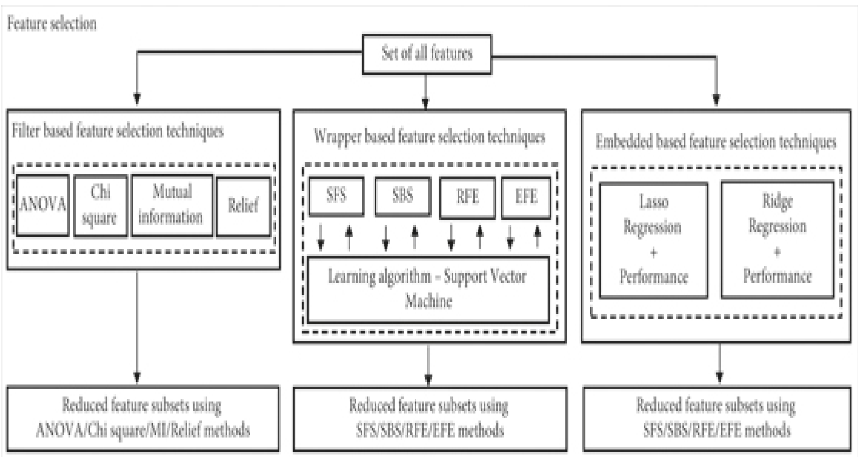


Figure:2 Feature selection

- **Feature selection using filter method:**

Filter methods for feature selection are indeed a valuable tool in machine learning and data analysis. They are particularly useful because they don't rely on any specific learning algorithm and instead assess features based on their intrinsic statistical properties. Here's a breakdown of how filter methods typically work: Independence Assessment: Filter methods evaluate the relationship between each independent feature (input attribute) and the dependent feature (target attribute). This assessment is done using statistical measures like correlation, mutual information, or chi-square tests, depending on the nature of the data.

- **Ranking or Scoring**

Once the independent assessment is complete, features are ranked or scored based on their statistical Once the independence assessment is complete, features are ranked or scored based on their statistical significance or relevance to the target variable. Features that exhibit strong relationships or high performance information gain with the target variable are typically ranked higher.

- **Threshold or Selection:**

A threshold may be set to select the top-ranking features or those above a certain score. Alternatively, all features above a certain threshold may be retained.

- **Validation and Fine Tunning:**

It's essential to validate the selected features to ensure they contribute meaningfully to the predictive performance of the model. This validation can involve techniques like cross-validation or testing the model's performance with and without the selected features.

- **Iterative process**

Feature selection using filter methods can be an iterative process, where different statistical measures are applied, and thresholds are adjusted to achieve the desired balance between feature relevance and model simplicity.

Common statistical method used in filter based features selection include:

- **Correlation Coefficient:**

Measures the strength and direction of the linear relationship between two variables. Features with high correlation to the target variable are considered relevant.

- **Mutual Information:**

Measures the amount of information obtained about one variable through the other variable. It quantifies the mutual dependence between variables.

- **Chi-square Test:**

Used for categorial variables, it tests the independence between variables. Its commonly used in features selection for classification tasks with categorial predictors.

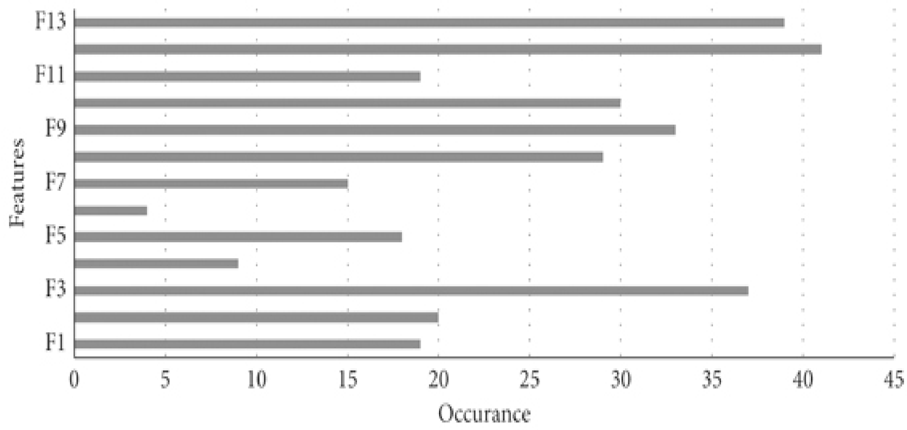


Figure: 3 Feature Selection Using Wrapper Methods

- **Feature Selection Using Wrapper Methods:**

Wrapper methods for feature selection involve selecting a subset of features by evaluating the performance of a machine learning algorithm trained on different combinations of features[12]. Here's a breakdown of the four main techniques mentioned:

- **Forward Feature Selection:**

Process: Start with an empty set of features. Iteratively add one feature at a time, selecting the one that maximize a chosen performance metric (e.g., accuracy, F1 score) on the validation set.
Advantages: It's computationally efficient compared to exhaustive search methods because it evaluates only one feature at a time.
Disadvantages: May not find the optimal feature subset due to its greedy nature.

- **Backward Feature Elimination:**

Process: Start with all features. Iteratively remove one feature at a time, choosing the one whose removal causes the least decrease in performance on the validation set.
Advantages: Guarantees convergence to the optimal subset when used with a suitable performance metric.
Disadvantages: Can be computationally expensive, especially with a large number of features.

- **Recursive Feature Elimination:**

Process: Train the model on all features and rank the importance of each feature based on a chosen criterion (e.g., feature weights, feature importance scores). Then, recursively remove the least important feature(s) until the desired number of features is reached.
Process*:* Train the model on all features and rank the importance of each feature based on a chosen criterion (e.g., feature weights, feature importance scores). Then, recursively remove the least important feature(s) until the desired number of features is reached.
Advantages*:* Automatically selects the optimal number of features based on the stopping criterion.
Disadvantages: Computationally expensive, especially for models that require retraining on subsets of features.

- **Exhaustive Feature Selection***:*

Process: Evaluate the performance of the learning algorithm on all possible feature subsets. This method requires evaluating all subsets of features, making it computationally expensive for large feature sets.
Advantages*:* Guarantees finding the optimal feature subset when using a suitable performance metric.
Disadvantages*:* Computationally infeasible for datasets with a large number of features due to the exponential growth in the number of subsets.

- **Feature Selection Using Embedded Methods:**

Embedded methods, as mentioned, integrate feature selection into the model training process. This means that feature selection is performed simultaneously with the model training, rather than as a separate step. Embedded methods are often specific to particular machine learning algorithms and are thus embedded within them. These methods work by evaluating the importance of features during the training process and making decisions on which features to retain or discard based on this evaluation[13]

Advantages of Embedded Methods:

Efficiency *:* Embedded methods typically require less computational resources compared to wrapper methods because feature selection is integrated directly into the model training process. This can make them more efficient, especially when dealing with large datasets or complex models.

Automatic Feature Selection:

Since feature selection is embedded within the model training process, the algorithm automatically selects the most relevant features based on their importance for the task at hand. This can lead to more streamlined and optimized models without the need for manual intervention.

Optimal Subset Generation:

Embedded methods aim to generate an optimal subset of features during the model training process. By integrating feature selection with model development, these methods can potentially lead to better-performing models by focusing on the most relevant features for the task.

- **Examples of Embedded Methods:**

Lasso Regression:

Lasso (Least Absolute Shrinkage and Selection Operator) regression is a linear regression technique that performs both regularisation and feature selection. It penalises the absolute size of the regression coefficients, effectively shrinking some coefficients to zero, thus performing feature selection.

- **Random Forest Feature Importance:**

In a random forest algorithm, feature importance can be computed based on how much each feature decreases impurity across all decision trees in the forest. Features with higher importance scores are considered more relevant, and the algorithm can automatically select features based on these scores.

- **Gradient Boosting Feature Importance**

Similar to random forests, gradient boosting algorithms can also compute feature importance scores based on how often features are used in decision trees during the boosting process. Features with higher importance scores are retained, while less important features may be discarded.

3 Proposed work

Our goal is to create predictions about heart disease. Cardiovascular disease is a major health problem worldwide, so the development of tools that can help predict cardiovascular disease appears to have a major impact on health. Our model will use machine learning techniques to analyse variables such as age, gender, blood pressure, cholesterol levels, and lifestyle factors such as smoking and counting. By incorporating these changes into our model, it will be able to predict a person's future risk of heart disease. First, we will collect a database containing information on people with and without a diagnosis of heart disease. This configuration file will form the basis for planning our actions. We will clean and prioritize the message to remove invalid or unnecessary information and then separate it into planning and testing phases. We will then choose the appropriate machine learning to prepare for our actions. This may include techniques such as iterative computation, selection trees, or support vector machines. We will prepare the program to improve its ability to accurately predict heart disease by using training data and changing its parameters. Overall, our goal is to provide effective and reliable predictions of heart disease that help in early detection and prevention of the disease, ultimately leading to better health outcomes for people at risk of heart disease.

- **There are several different methods:**

Discriminators*:*

These are the calculations or formulas that make the prediction itself. Examples include decision trees, inverse vector machines, k-nearest neighbors, etc. takes place.

Identification or reduction of the process:

This strategy involves selecting one of the main points or reducing the rest of the important places that have recently contributed to the distributed process. Techniques such as principal component analysis (PCA) involve returning spatial significance or importance to this group.

Integrated system:

Good clothing combines many modeling systems to improve forecasting. For example, Arbitrary Woodland, Angular Booster Machine (GBM), AdaBoost, etc. This strategy works by maintaining the expectations of different processes.

Datasets:

Common Traits in Categorical/Integer/Real Datasets:

A. Age: Age of the patient.
B. Sex: Sex of the patient.
C. Chest Torment: Portrayal or seriousness of chest pain.
D. Resting Blood Weight: Blood weight of the quiet at rest.
E. Cholesterol: Cholesterol levels in the patient's blood.
F. Fasting Blood Sugar: Blood sugar levels in the patient's blood after fasting
G. Resting ECG Comes about: Comes about of electrocardiogram tests performed at rest.
H. Greatest Heart Rate: Greatest heart rate accomplished amid exercise.
I. Exercise-induced Angina: Nearness or nonappearance of angina (chest torment) actuated by exercise.
J. Slope of Top Work out: Incline of ST fragment amid crest exercise.

Features of ECG Signals:

ECG signals consist of highlights from two major spaces: time and frequency.

Time Space Features:

1. Cruel RR Interims: Cruel of all RR interims (time between progressive heartbeats).
2. Standard Deviation RR: Standard deviation of RR intervals.
3. Square of the Cruel of the Whole of the Distinction between Adjoining **Interims:** A measure of heart rate variability.

Frequency Space Features*:*

4. Control of Low and High-Frequency Groups: Control ghostly thickness in low-frequency and high-frequency bands.

Nonlinear Elements Measures*:* Nonlinear elements measures may shift depending on the strategies connected or proposed in the analysis. These measures seem to incorporate different complexity measures, entropy measures, fractal dimensions, and other nonlinear strategies to capture the complexity and elements of heart rate variability and ECG signals.

4 Related work

In this segment, the essential revelations of the chosen papers have been pondered after a thorough analysis. It gives a summation of the techniques laid out in the chosen papers. The inquiries about papers are categorized based on the classification strategies utilized. A few of these papers have connected their techniques to an assortment of datasets counting Diabetes, Heart Infection, Breast Cancer, Liver Infection, and Hepatitis datasets. However, for the purposes of this overview, papers centering on heart infection datasets have been organised. These datasets incorporate UCI, PHP, BIDMC CHF dataset, PTB Demonstrative ECG, among others. The taking after strategies are found to be most prevalently utilized for heart infection and HF classification.

Support Vector Machine (SVM) :

Sung and Yuan Lee utilise Hereditary Calculation (GA) for Feature choice and SVM for classification. This technique approves its viability. When SVM is connected without GA, it outclasses two other methods in writing by producing an precision of 96.38%, when GA is connected the accuracy is progressed by 3.14%.. makes utilisation of Vital Component Investigation (PCA) for dimensionality decrease and at that point applies SVM for classification purpose.

Neural Networks:

The neural network serves as the core component for CAD diagnosis. It takes inputs (possibly patient data, medical history, diagnostic tests, etc.) and processes them through layers of interconnected nodes (neurons) to produce an output, which could be a binary classification indicating the presence or absence of CAD.

Decision Trees :

The utilisation of Dual Tree Complex Wavelet Transform (DTCWT) for automatically distinguishing congestive heart failure (CHF) from normal electrocardiogram (ECG) signals is a promising approach in biomedical signal processing. By employing DTCWT, the ECG signals are decomposed into their frequency components, enabling a detailed analysis of the signal characteristics.In this proposed methodology, ECG segments of 2 seconds are utilized, allowing for rapid processing and decision-making. The study employs five different techniques for feature ranking, which likely involve statistical measures or signal processing algorithms to extract relevant features from the transformed ECG signals. These features are then inputted into classification algorithms such as k-Nearest Neighbors (kNN) and decision trees.[6]

KNN:

In the first study, various classifiers including k-nearest neighbour (kNN), decision trees (DT), and support vector machines (SVM) were used on a dataset with clinically important features. Among these classifiers, SVM achieved the highest accuracy. In the second study, non-linear features were used as input to kNN and SVM classifiers. The accuracy achieved by kNN was 92.11%, while SVM achieved an accuracy of 96.68%. These studies demonstrate the importance of feature selection and the choice of classifier in achieving high accuracy in classification tasks. SVM seems to perform consistently well across different feature sets in both studies.[7]

Random Forest:

These modifications are aimed at improving the accuracy of the Random Forest model in predicting outcomes related to survival, such as in medical prognosis or risk assessment scenarios. By enhancing the split rule and stopping criteria, the Improved Random Forest may better capture the underlying patterns in the data, resulting in more accurate predictions.[8]

5 Discussion

A heart disease prediction model is like a smart tool that uses information about a person's health and lifestyle to guess the likelihood of them developing a certain condition—in this case, heart disease. It's kind of like a weather forecast, but instead of predicting rain or sunshine, it predicts the likelihood of someone getting sick.The model looks at various factors that can contribute to heart disease, such as age, gender, blood pressure, cholesterol levels, whether the person smokes, and so on. It analyses how these factors relate to each other and to the likelihood of heart disease. Think of it as putting together puzzle pieces to see the bigger picture. Early detection of heart disease is crucial because it allows people to make lifestyle changes, like eating healthier or exercising more, which can reduce the risk of heart problems. It also helps doctors provide appropriate treatments or medications to manage the condition effectively[9].We'll gather data from different sources, like medical records or surveys, and use computer programs to analyZe it. Then, we'll train our prediction model using this data, teaching it how to recognize patterns that indicate a higher risk of heart disease. Once the model is trained, we can test it to see how accurate it is at predicting heart disease in new cases.

Heart failure (HF) stands as one of the leading causes of mortality worldwide, underscoring the critical importance of accurate risk prediction to prevent and manage it effectively. Timely identification of congestive heart failure (CHF) is particularly crucial in avoiding life-threatening events. Accuracy in medical diagnosis is paramount, given its direct impact on patient outcomes. Extensive research has been conducted on disease classification and prediction using machine learning (ML) techniques, although consensus on the optimal technique or classifier for specific datasets remains elusive. Nonetheless, it is established that feature selection and reduction techniques enhance classifier accuracy and reliability. Furthermore, ensemble classifiers have demonstrated improved classification accuracy, further highlighting their potential in medical applications [2].

6 Conclusion

Our minor project focused on developing a heart disease prediction model aimed at assisting in early detection and prevention of cardiovascular issues. Through the utilisation of machine learning algorithms and a dataset comprising various health parameters and medical history, we aimed to create a tool that could potentially aid healthcare professionals in identifying individuals at higher risk of heart disease. Our model underwent rigorous testing and evaluation to ensure its accuracy and reliability. By analysing factors such as age, gender, blood pressure, cholesterol levels, and lifestyle habits, our model was able to generate predictions with a significant degree of accuracy. This predictive capability holds promise in potentially reducing the incidence of heart disease by allowing for targeted interventions and lifestyle modifications. One of the key strengths of our model is its simplicity and ease of use. Healthcare professionals can input patient data into the system, and the model will provide an instant assessment of the individual's risk of developing heart disease. This user-friendly interface makes it accessible to a wide range of medical personnel, from general practitioners to specialists, thereby increasing its utility and impact within the healthcare community.

Machine learning-based solutions are widely used in healthcare to analyse patient data, predict diseases and recommend possible treatments. Since there are many types of machine learning available today, it is important to identify effective and accurate methods, especially in important areas such as treatment. A comparison of various machine learning algorithms used in cardiovascular disease prediction was made. The five most commonly used methods: KNN, logistic regression, naive Bayes, decision trees, and random forests are discussed and compared to determine the most appropriate method for predicting heart disease. Many previous studies and studies on the prediction of cardiovascular disease were identified and reviewed. Overall, the findings show that machine learning techniques have the potential to transform the healthcare industry and improve overall disease prediction and reporting treatment.

7 References

[1] Babu, S., Vivek, E., Famina, K., Fida, K., Aswathi, P., Shanid, M., Hena, M.: Heart disease diagnosis using data mining technique. In: *2017 International conference, Communication and Aerospace Technology (ICECA).* vol. 1, pp.750–753. IEEE (2017).

[2] Cai, J., Luo, J., Wang, S., Yang, S.: Feature selection in machine learning: A newperspective. Neurocomputing 300, 70–79 (2018).

[3] . sandeep K. M, jain, M. A Data Driven Deep Neural Network Model for Identifying both Covid-19 Disease along with Potential Pandemic Hotspots. *https://ijisae.org/index.php/IJISAE/article/view/5065/3772* volume 12, 19s, 284–295.

[4] Dangare, C.S., Apte, S.S.: Improved study of heart disease prediction system using data mining classification techniques. International Journal of Computer Applica-tions 47(10), 44–48 (2012).

[5] Fang, X., Hodge, B.M., Du, E., Zhang, N., Li, F.: Modelling wind power spatial-temporal correlation in multi-interval optimal power flow: A sparse correlation matrix approach. Applied energy 230, 531–539 (2018).

[6] Gavhane, A., Kokkula, G., Pandya, I., Devadkar, K.: Prediction of heart diseaseusing machine learning. In: 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA). pp. 1275–1278. IEEE (2018).

[7] Hasan, S., Mamun, M., Uddin, M., Hossain, M.: Comparative analysis of clas-sification approaches for heart disease prediction. In: 2018 International Confer-ence on Computer, Communication, Chemical, Material and Electronic Engineer-ing (IC4ME2). pp. 1–4. IEEE (2018).

[8] S.K. Mathariya et al. Acquiring the Ability to Identifying Covid19 using Deep CNN from Impulse Noise in Chest X-Ray Pictures: 2024. *https://ijisae.org/index.php/IJISAE/article/view/5429*. Accessed: 2024-04- pp 347-353(2024)

[9] Jenzi, I., Priyanka, P., Alli, P.: A reliable classifier model using data mining approach for heart disease prediction. International Journal of Advanced Research in Computer Science and Software Engineering 3(3) (2013).

[10] Kalaiselvi, C.: Diagnosing of heart diseases using average k-nearest neighbor algorithm of data mining. In: 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom). pp. 3099–3103. IEEE (2016).

[11] Lopez, A.D., Mathers, C.D., Ezzati, M., Jamison, D.T., Murray, C.J.: Global andregional burden of disease and risk factors, 2001: systematic analysis of population health data. The Lancet 367(9524), 1747–1757 (2016).

[12] Masethe, H.D., Masethe, M.A.: Prediction of heart disease using classification al-gorithms. In: Proceedings of the world Congress on Engineering and computer science. vol. 2, pp. 22–24 (2014).

[13] Organization, W.H.: The world health report 2002: reducing risks, promoting healthy life. World Health Organization (2016)12. Organization, W.H., of Canada, P.H.A., of Canada, C.P.H.A.: Preventing chronic diseases: a vital investment World Health Organization (2015)13.

[14] Patel, S.B., Yadav, P.K., Shukla, D.: Predict the diagnosis of heart disease patients using classification mining techniques. IOSR Journal of Agriculture and Veterinary Science (IOSR-JAVS) 4(2), 61–64 (2013).

[15] Peter, T.J., Somasundaram, K.: An empirical study on prediction of heart disease using classification data mining techniques. In: IEEE-International conference on advances in engineering, science and management (ICAESM-2012). pp. 514–518.IEEE (2012).

[16] priyanka, N., RaviKumar, P.: Usage of data mining techniques in predicting the heart diseases—na¨ıve bayes & decision tree. In: 2017 International Conference on Circuit, Power and Computing Technologies (ICCPCT). pp. 1–7. IEEE (2017).

[17] IEEE (2016)9. Lopez, A.D., Mathers, C.D., Ezzati, M., Jamison, D.T., Murray, C.J.: Global andregional burden of disease and risk factors, 2001: systematic analysis of population health data. The Lancet 367(9524), 1747–1757 (2016)

[18] Masethe, H.D., Masethe, M.A.: Prediction of heart disease using classification al-gorithms. In: Proceedings of the world Congress on Engineering and computer science. vol. 2, pp. 22–24 (2014)11.

[19] [4]Organization, W.H.: The world health report 2002: reducing risks, promoting healthy life. World Health Organization (2016)12. Organization, W.H., of Canada, P.H.A., of Canada, C.P.H.A.: Preventing chronic diseases: a vital investment.

[20] [5]World Health Organization (2015)13. Patel, S.B., Yadav, P.K., Shukla, D.: Predict the diagnosis of heart disease patientsusing classification mining techniques. IOSR Journal of Agriculture and VeterinaryScience (IOSR-JAVS) 4(2), 61–64 (2013)14.

[21] [6]Peter, T.J., Somasundaram, K.: An empirical study on prediction of heart diseaseusing classification data mining techniques. In: IEEE-International conference onadvances in engineering, science and management (ICAESM-2012). pp. 514–518.IEEE (2012)15.

[22] Priyanka, N., RaviKumar, P.: Usage of data mining techniques in predicting theheart diseases—na¨ıve bayes & decision tree. In: 2017 International Conference onCircuit, Power and Computing Technologies (ICCPCT). pp. 1–7. IEEE (2017)16.

[23] M Momeny, AA Neshat, MA Hussain, S Kia, "Learning-to-augment strategy using noisy and denoised data: Improving generalizability of deep CNN for the detection of COVID-19 in X-ray images", Computers in Biology and Medicine, Elsevier, 2022, vol.136, 104704.

[24] MF Aslan, MF Unlersen, K Sabanci, A Durdu, "CNN-based transfer learning–BiLSTM [22]NN Das, N Kumar, M Kaur, V Kumar, D Singh, "Automated deep transfer learning-based approach for detection of COVID-19 infection in chest X-rays", IBRM, Elsevier, 2022, vol.43, no.2, pp.114-119.

[25]]D. A. Zebari, A. M. Abdulazeez, D. Q. Zeebaree and M. S. Salih, "A Fusion Scheme of Texture Features for COVID-19 Detection of CT Scan Images," 2020 International Conference on Advanced Science and Engineering (ICOASE), 2020, pp. 1-6.

[26] .El Gannour, S. Hamida, B. Cherradi, A. Raihani And H. Moujahid, "Performance Evaluation of Transfer Learning Technique for Automatic Detection of Patients with COVID-19 on X-Ray Images," 2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), 2020, pp. 1-6