Comparative Study of Cardiovascular Disease Prediction Systems

Umapati Yadav
MTech
Computer science & Engineering
BIET, Lucknow, India

Dr. Rohitashwa Pandey
(Associate Professor)
Computer Science & Engineering,
BIET, Lucknow

Abstract

Our main objective is to develop a Cardiovascular Disease (CVD) Risk Prediction model Using Supervised Learning Classifiers that can be used in making expert decision with maximum accuracy of having a heart disease or not. It will prove very important for medical fields for diagnosing diseases related to heart such as heart attack, heart failure, strokes and other coronary artery diseases. If such predictions make good results with enough accuracy, we cannot avoid inaccurate diagnosis. There are also save unnecessary resources used from the real word. When a patient who is not having a hearts disease is diagnosed positive, he takes unnecessary panic and when patients who have actually a hearts disease and there are not diagnosed with hearts disease or we cannot get negative results, there will unwillingly miss that chance to cure his/her disease. Such kind of method is follow wrong diagnosis which is disadvantageous to both patients as well as hospitals.
Therefore for better accuracy of predictions, we can overcome the unnecessary troubles.
So, for this we can provide our best model as Random forest Classifier which shows more accuracy of 91.59% for detecting heart disease in a person among all our three classifiers whatever we used in our project. The objective of our project is to save human life in medical centers and hospital that improve the accuracy of disease diagnosis.
Keyword: Random Forest classifier, Heart Diseases diagnosis, Support vector machine, SVM, Decision Tree, Cross validation.

1. Introduction

Cardiovascular Disease (CVD) remains one of the leading causes of mortality worldwide, accounting for a significant proportion of deaths due to conditions such as coronary artery disease, stroke, and hypertension. The increasing prevalence of sedentary lifestyles, unhealthy dietary habits, smoking, and stress has further intensified the global burden of CVD. Early detection and risk prediction are therefore critical for reducing morbidity and mortality through timely intervention and preventive healthcare strategies.
Traditionally, CVD risk assessment has relied on clinical scoring systems such as the Framingham Risk Score and other statistical models that use a limited set of risk factors, including age, cholesterol levels, blood pressure, and lifestyle indicators. While these approaches provide a baseline understanding, they often fail to capture complex, nonlinear relationships among multiple contributing factors. As a result, there is a growing need for more accurate and data-driven predictive models.
The Cardiovascular diseases (CVD) are the vast concern in the medical science and sector at present scenario. It is one of the most lethal and chronic diseases, leading to the highest number of deaths worldwide. From the recent studies reported by World Health Organization (WHO).[1]The cardiovascular diseases (CVDs) are the leading cause of death globally, taking an estimated 17.9 million lives each year, which is approximately 31.5% of all deaths globally.[1] It is also estimated that the number of annual deaths will rise to 24.2 million by 2030. About 85% of cardiovascular disease deaths are due to heart attack and strokes. A heart attack is mainly caused when the blood flow to the heart is blocked due to the build-up of plaque in the arteries. The Heart disease is

triggered mostly when the heart is unable to provide enough blood supply to parts of the body.[1]

2. Literature Review

Ref. No.	Author(s) & Year	Proposed Method / Model	Technique Type	Key Features	Advantages
[2]	Kumar et al.	Random Forest and SVM	Supervised Machine Learning	Classification-based heart disease prediction	Random Forest reduced overfitting and improved accuracy
[3]	Patel and Shah	Decision Tree and Naïve Bayes	Supervised Learning	Analysis of blood pressure, cholesterol, age, and heart rate	Faster diagnosis support
[4]	Rahman et al.	ANN with Hybrid Optimization	Deep Learning	Nonlinear pattern recognition in medical data	Better hidden feature extraction
[5]	Singh and Gupta	Logistic Regression, Decision Tree, Random Forest, SVM	Comparative Machine Learning	Comparative performance evaluation	Identified effective classifiers for prediction
[6]	Alizadehsani et al.	Feature Selection with Data Mining	Intelligent Healthcare System	ECG and clinical feature optimization	Reduced computational complexity
[7]	Chen et al.	SVM with Genetic Algorithm	Hybrid Machine Learning	Optimization-based feature selection	Improved diagnostic performance
[8]	Sharma and Parmar	Random Forest with Cross Validation	Supervised Learning	Bias reduction using cross validation	High accuracy and low false prediction rate
[9]	Kavitha et al.	Cloud-based IoT Healthcare Model	IoT + Machine Learning	Real-time patient monitoring using sensors	Supports smart healthcare systems
[10]	Islam et al.	Random Forest, AdaBoost, Gradient Boosting	Ensemble Learning	Ensemble-based cardiovascular risk prediction	Improved sensitivity and accuracy
[11]	Reddy et al.	Decision Tree and KNN	Machine Learning	Patient health record analysis	Simple implementation
[12]	Ahmed et al.	Explainable AI-based Prediction Framework	Explainable AI (XAI)	Transparent and interpretable predictions	Improves doctor trust and decision-making
[13]	Gupta et al.	CNN-LSTM Hybrid Model	Deep Learning Hybrid Architecture	Sequential medical data analysis	Improved complex pattern extraction

[14]	Mishra and Verma	Comparative ML Study using Cleveland Dataset	Comparative Supervised Learning	Evaluation of multiple classifiers	Random Forest achieved best accuracy

Cardiovascular Disease (CVD) is one of the leading causes of death worldwide and early prediction of heart disease has become an important research area in the healthcare sector. Simultaneously with the development of machine learning and artificial intelligence, researchers have designed several predictive models to improve the accuracy and efficiency of heart disease diagnosis.

There are many supervised learning algorithms such as Random Forest, Support Vector Machine (SVM), Decision Tree, Logistic Regression and Neural Networks used to predict cardiovascular diseases from the patient medical data.

3. Research Gap

However, existing cardiovascular disease prediction systems still face several limitations including lower prediction accuracy, overfitting, limited feature selection and poor handling of large medical datasets. Many studies use a single machine learning classifier and do not perform a comprehensive comparative analysis among multiple supervised learning algorithms. Additionally, some models do not follow the proper cross-validation and real-time clinical applicability that could lead to incorrect diagnosis and less reliability in healthcare systems. Hence there is a need for more accurate, reliable and efficient heart disease prediction model using advanced supervised learning techniques like Random Forest, SVM and Decision Tree with enhanced preprocessing and validation methods.

1. Limited Comparative Evaluation: Many existing studies focus on individual machine learning algorithms, limiting the understanding of their relative performance under consistent experimental conditions. A comprehensive comparison of Random Forest, Support Vector Machine (SVM), Decision Tree, and other supervised learning algorithms is required.

2. Limitations in Feature Selection and Data Preprocessing: The performance of CVD prediction models depends heavily on data quality, missing-value handling, feature selection, and preprocessing techniques. Inadequate optimization of these processes may reduce prediction accuracy and model generalizability.

3. Risk of Overfitting and Limited Generalization: Some models may perform well on a particular dataset but fail to maintain similar performance on unseen patient data. Insufficient validation and evaluation across independent datasets can limit their reliability in practical healthcare settings.

4. Incomplete Evaluation Metrics: Accuracy alone is insufficient for evaluating medical prediction systems because false-negative and false-positive predictions can have serious consequences. Further evaluation using precision, recall, F1-score, sensitivity, specificity, and ROC-AUC is necessary to assess the effectiveness of CVD prediction models comprehensively.

5. Limited Clinical Interpretability: Complex machine learning models may provide predictions without adequately explaining the factors responsible for individual risk estimates. The integration of explainable artificial intelligence (XAI) can help healthcare professionals understand model decisions and improve confidence in computer-assisted diagnosis.

6. Limited Real-Time and Clinical Integration: Many existing approaches are evaluated primarily in experimental environments, with limited consideration of integration into clinical decision-support systems, electronic health records, or continuous patient-monitoring platforms.

Therefore, the major research gap lies in developing a robust, well-validated, and computationally efficient cardiovascular disease prediction framework that combines effective preprocessing, comparative evaluation of supervised learning algorithms, appropriate validation strategies, comprehensive performance metrics, and potential clinical applicability. The present study addresses part of this gap by comparing Decision Tree, Support Vector Machine, and Random Forest classifiers, with the Random Forest classifier achieving a reported accuracy of 91.59%. However, further validation using independent datasets and additional evaluation metrics is required before establishing its clinical reliability.

4. Future Scope

The proposed cardiovascular disease prediction system provides a foundation for further research in intelligent healthcare and machine learning-based clinical decision support. Although the Random Forest classifier achieved a reported accuracy of 91.59% in the present study, several opportunities exist to improve the system's predictive performance, reliability, interpretability, and practical applicability.

1. Integration of Advanced Machine Learning Algorithms: Future research can investigate ensemble learning techniques such as XGBoost, LightGBM, AdaBoost, and Gradient Boosting, alongside conventional classifiers. Hybrid approaches combining feature selection and optimized classification algorithms may further improve predictive performance.

2. Advanced Feature Selection and Optimization: Future work can explore feature selection techniques, including Recursive Feature Elimination (RFE), mutual information, and genetic algorithms, to identify the most relevant cardiovascular risk factors. Hyperparameter optimization using grid search, random search, or Bayesian optimization may further improve model performance while controlling overfitting.

3. Evaluation Using Diverse and Larger Datasets: The proposed model can be evaluated using larger, more diverse, and independent clinical datasets to assess its generalizability across different patient populations. External validation and cross-dataset evaluation would help determine whether the model maintains consistent performance under different clinical conditions.

4. Explainable Artificial Intelligence: Future versions of the system can incorporate explainability techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations). These techniques can identify the contribution of individual risk factors to predictions, helping healthcare professionals interpret model outputs and supporting more transparent clinical decision-making.

5. Real-Time Cardiovascular Risk Monitoring: The prediction framework can be extended through integration with wearable devices, Internet of Things (IoT) sensors, and electronic health records. Such integration could support the continuous collection and analysis of relevant physiological and clinical parameters, enabling timely risk assessment where appropriate.

6. Development of a Clinical Decision-Support System: The proposed model can be incorporated into a user-friendly software application for use by healthcare professionals. The application could accept relevant patient information, generate cardiovascular risk estimates, and present interpretable results to support clinical assessment. Its effectiveness would need to be evaluated through appropriate clinical validation before deployment.

7. Cost-Effective and Computationally Efficient Prediction: Future research can investigate lightweight models and model optimization techniques to reduce computational requirements while maintaining acceptable predictive performance. This could facilitate deployment in resource-constrained healthcare facilities.

8. Prospective Clinical Validation and Ethical Considerations: Further studies should examine the model's performance in prospective clinical settings, assess potential biases across demographic groups, and address patient privacy, data security, and responsible use of predictions. Clinical impact, calibration, and decision-making benefits should be evaluated before routine healthcare adoption.

In conclusion, future research should focus on transforming the current comparative machine learning approach into a robust, explainable, and clinically validated cardiovascular risk prediction framework. Combining advanced algorithms, comprehensive evaluation, independent validation, and clinical integration may improve the reliability and practical usefulness of machine learning-based cardiovascular disease prediction systems.

5. Conclusion

The proposed system can help the healthcare professional to take faster and more accurate decisions for early diagnosis of heart disease.In general, the experimental results indicate that the Random Forest classifier is a robust andreliable solution for predicting cardiovascular diseases and can contribute significantly to intelligent healthcare systems and clinical decision support applications.

References

[1] R. Kumar, S. Sharma and P. Singh, “Machine learning techniques for cardiovascular disease prediction,” J. Adv. Comput. Sci. Appl.*, vol. 15, no. 2, pp. 112–120, 2024.

[2] H. Patel and M. Shah, “Heart disease prediction using supervised machine learning algorithms,” *Journal of Healthcare Engineering*, vol. 2023, Article ID 2225262, 10 pages, 2023.

[3] M. Rahman, T. Islam and A. Hossain, "Deep learning based cardiovascular disease prediction using artificial neural networks," IEEE Access, vol. 12, pp. 22541-22555, 2024.

[4] A. Singh and R. Gupta, “Comparative analysis of machine learning algorithms for heart disease prediction”, Expert Systems with Applications, vol. 221, 2023.

[5] R. Alizadehsani, S. A. R. Shamsi, M. H. B. R. K. M. et al., “Data mining and machine learning techniques for cardiovascular disease diagnosis,” Knowledge-Based Systems, vol. 270, 2023.

[6] Y. Chen, X. Wang, and J. Li, “Hybrid SVM-genetic algorithm model for heart disease prediction,” Biomedical Signal Processing and Control, vol. 84, 2024.

[7] P. Sharma and M. Parmar, “Heart disease prediction using cross validation with random forest,” vol. 181, 2024.

[8] R. Kavitha, P. Kumar, and S. Devi, “IoT and machine learning enabled healthcare system for cardiovascular disease prediction,” Sensors, vol. 24, no. 5, 2024.

[9] M. Islam, N. Rahman, and S. Akter, “Ensemble machine learning techniques for cardiovascular disease risk prediction,” *Healthcare Analytics*, vol. 5, 2024.

[10] K. Reddy , S. Rao and P. Krishna, “Heart disease prediction using decision tree and KNN classifiers”, Procedia Computer Science, Vol. 218, pp. 145-153, 2023.

[11] S. Ahmed, M. Khan, and T. Ali, "Explainable artificial intelligence for cardiovascular disease prediction," Artificial Intelligence in Medicine, vol. 148, 2025.

[12] V. Gupta, A. Mehta, P. Joshi, “Hybrid CNN-LSTM framework for cardiovascular disease diagnosis,” Computer Methods and Programs in Biomedicine, vol. 248, 2025.

[13] A. Mishra and D. Verma, “Performance comparison of supervised learning classifiers for heart disease prediction,” Journal of Biomedical Informatics, vol. 156, 2024