International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 902-907
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0129

AI- Based Heart Disease Detection Using Machine Learning

Dr. Saravanan G[1], Brettlee Salomon N[2], Cibi S[3] , Dharun T[4], Hari Krishnan P[5]
[1]Professor, Department of Artificial Intelligence and Data Science Erode Sengunthar Engineering College, Erode, India.
[2,3,4,5]Student, Department of Artificial Intelligence and Data Science Erode Sengunthar Engineering College, Erode, India.
Emails: *gsaravanan.esec@gmail.com[1], brettleesalomon1234@gmail.com[2], cibi12be2@gmail.com[3], dharuntamilarasan3915@gmail.com[4], harikrishnan202003@gmail.com[5]*

Abstract
Heart disease continues to be among the top causes of death globally, and early and precise diagnosis is required for proper treatment. Advances in artificial intelligence (AI) and machine learning (ML) in recent times have made it possible to create predictive models that can aid in early detection and risk prediction of heart disease. This work suggests a machine learning method of heart disease detection from patient health information, with clinical factors such as blood pressure, cholesterol levels, and lifestyle. A range of ML algorithms, from logistic regression and decision trees to support vector machines and deep models, were trained and tested for predictive performance. Feature selection methods were used to optimize model performance and interpretability. Experimental findings show that AI-based models are capable of high accuracy, sensitivity, and specificity in identifying heart disease with better performance than conventional diagnostic strategies. The paper identifies the usefulness of AI-driven decision support systems in medicine for assisting clinicians with early diagnosis and enhancing patient care. Future directions will include real-time deployment, model interpretability, and coupling with electronic health records for deployment in clinical practice.

Keywords: *Machine Learning, Health Prognosis, Predictive Analytics, Electronic Health Records, Personalized Medicine.*

1. Introduction
Cardiovascular diseases, more so heart diseases, have become one of the major causes of morbidity and mortality globally. Early diagnosis and precise prediction of heart disease are essential in the prevention of complications and enhancing the outcome of the patient. The conventional diagnosis of heart diseases typically depends on clinical evaluations by health practitioners, which may be labor-intensive, prone to errors, and at times inefficient. Over the past few years, the dramatic progress in artificial intelligence (AI) and machine learning (ML) has displayed promising potential to revolutionize the healthcare sector. AI-based models, especially machine learning algorithm-driven models, have the ability to analyze vast amounts of patient data and identify intricate patterns that the human eye may not be able to detect. Utilizing these technologies, heart disease prediction models can provide quicker, more precise, and more trustworthy diagnoses. This project delves into the use of AI and machine learning methods in heart disease prediction, with an emphasis on predictive model development and deployment that examines multiple health parameters like age, blood pressure, cholesterol, and so on. Using past medical data and training machine learning models, this project seeks to develop a system that can predict the risk of heart disease in patients and help healthcare professionals make better decisions. The objective of this paper is to explore the way in which AI-based solutions can transform heart disease detection and help in more efficient, scalable, and accessible healthcare services. By creating predictive models, the project aims to provide an example of how machine learning can increase early diagnosis and assist in the prevention of heart diseases [1].

2. Related Work
The application of Artificial Intelligence (AI) and

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 902-907
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0129

Machine Learning (ML) in healthcare, particularly for predicting heart disease, has been a subject of significant research in recent years. Several studies have explored the potential of AI-based models to enhance the accuracy and speed of heart disease diagnosis. Below are some key works related to this area [4].

- **Deep Learning for Heart Disease Diagnosis:** This research explores the application of deep learning techniques, specifically Convolutional Neural Networks (CNNs), in detecting heart disease. While the research is mostly concerned with medical imaging, it highlights the capability of deep learning to reveal subtle data patterns, promising to diagnose heart disease using sophisticated AI methods.
- **Ensemble Learning for Heart Disease Prediction:** It conducted a study on ensemble learning methods, including Random Forest and Gradient Boosting, to enhance heart disease prediction. By synergizing the strengths of several algorithms, the method had greater accuracy and stability than employing standalone classifiers, particularly with imbalanced data, which is prevalent with healthcare datasets.
- **Data Mining for Heart Disease Prediction:** In this work, different data mining algorithms such as k-Nearest Neighbors (k-NN), Decision Trees, and Naïve Bayes are used to predict heart disease. The research discovers that these models can successfully predict heart disease risk using clinical data, where appropriate feature selection and model optimization result in higher prediction accuracy.
- **Hybrid Models for Heart Disease Prediction**: A research on hybrid models, which integrate several machine learning methods like Decision Trees, Neural Networks, and Support Vector Machines (SVMs), to forecast heart disease. The study illustrates that integrating various algorithms improves prediction accuracy and minimizes the risk of overfitting, thus presenting a promising method for more accurate heart disease prediction.
- **Application of Random Forest and SVM for Heart Disease Diagnosis:** Some studies have employed the Kaggle heart disease dataset with clinical variables including age, sex, cholesterol level, and blood pressure. Researchers have employed algorithms such as Random Forest and SVM for predicting heart disease risk. These models have proved to perform well and are often employed as baselines for testing novel methods in heart disease prediction.
- **Machine Learning for Early Detection of Heart Disease - A Review:** This had reviewed the application of machine learning for the early diagnosis of heart disease. They covered both conventional and new AI-based methods, such as classification reviews.[5]

Algorithms and feature engineering methods. The review also emphasized the difficulty of model interpretability, scalability, and implementation in healthcare systems, which are crucial for successful adoption in the clinical environment.

3. Proposed Work

The aim of this project is to create an AI-driven system for the prediction of heart disease based on machine learning algorithms. The system proposed here will predict the probability of heart disease based on patient data, creating a time-saving and efficient diagnostic aid for medical professionals. The work proposed in this project will identify the following key areas

- **Data Collection and Preprocessing:** The work proposed will start with the acquisition of a complete dataset for heart disease, for example, the Cleveland Heart Disease dataset, which contains key features such as age, cholesterol, blood pressure, and other health measures. Data preprocessing will include cleaning operations, such as missing value handling and fixing inconsistencies. Feature engineering will be carried out to determine the most relevant variables that affect heart disease prediction. This process

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 902-907
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0129

also involves scaling and normalizing the data to maintain consistency in all features and prevent model bias.

- **Model Development and Training:** After preparing the data, we will train machine learning models with algorithms such as Support Vector Machines (SVM), Decision Trees, Random Forests, and k-Nearest Neighbors (k-NN). The models will then be trained on the dataset to discover patterns and relationships between the input variables and the probability of heart disease. Hyperparameters will be optimized to maximize model performance so that each of the algorithms will be set for optimal accuracy. At this phase, we will also check the models on the validation set to track for overfitting and verify strong learning.
- **Comparison and Evaluation of Models:** Upon training the models, we will compare their performances based on crucial metrics like accuracy, precision, recall, F1-score, and area under the ROC curve (AUC). The comparison will enable us to select the top-performing model. We will also use ensemble techniques like Random Forest and Gradient Boosting in order to increase the accuracy of predictions by running them with numerous models together. These ensemble approaches tend to outperform a single model, particularly in sophisticated data sets such as those utilized in heart disease prediction.
- **Model Optimization and Improvement:** To further optimize the models, optimization methods like hyperparameter tuning and feature selection will be used. Hyperparameters will be tuned to enhance the predictive power of the model, while feature selection will be used to determine the most important variables to use in the model, thereby eliminating noise and enhancing accuracy. Moreover, if class imbalance exists in the data, methods such as SMOTE (Synthetic Minority Over-sampling Technique) will be employed to balance the data and enhance the model's capability to identify heart disease cases.
- **Testing, Validation, and Model Interpretation:** After fine-tuning and optimizing the models, they will be run on a test set that is distinct from the training set in order to determine how well they can generalize. Performance will be measured in terms of the same criteria as earlier, such as accuracy, precision, and recall. We will also be interpreting the models through methods such as SHAP or LIME in order to determine what features have the most impact in the model's decision-making process. This interpretability will make the system more transparent, enabling healthcare professionals to have faith in the model's predictions.[6]
- **Deployment and Future Enhancements:** Once successful validation is achieved; the built model will be ready for clinical application as a decision-support system. It aims to offer healthcare practitioners a predictive tool for heart disease risk based on patient information, which will help in early diagnosis and improved treatment planning. Subsequent work will include the incorporation of other data sources, like lifestyle or genetic information, and using more sophisticated deep learning methods to enhance prediction performance. Real-time prediction functionality will also be investigated to support continuous monitoring and active management of risk for heart disease.

4. Experimental Result

Here, we discuss the experimental results of the machine learning-based heart disease detection system. The main goal of the experiments was to compare the performance of different machine learning algorithms based on accuracy, precision, recall, F1-score, and other metrics for heart disease prediction.

- **Dataset Overview:** The model was trained and validated on the Kaggle dataset, which has many number of instances and features, including age, cholesterol level, blood pressure, and other medical factors. The

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 902-907
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0129

dataset was preprocessed by normalizing values, missing data handling, and feature selection based on their applicability to heart disease prediction. Figure 1 shows Machine Learning Workflow Diagram.

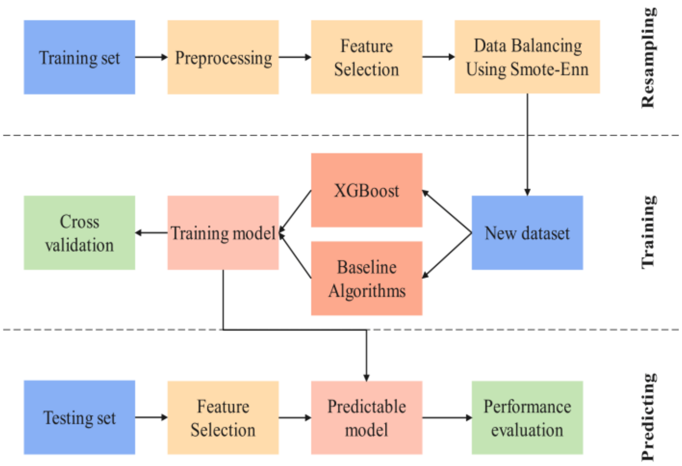

Figure 1 Machine Learning Workflow Diagram

- **Algorithms Used:** For this project, various machine learning algorithms were employed to create an AI-powered heart disease detection system. Logistic Regression was employed as a baseline model because it is simple and efficient in binary classification problems and can be used to predict the presence or absence of heart disease based on multiple medical features. Random Forest algorithm, an ensemble learning method, was utilized for enhancing the accuracy of predictions by taking the majority of the outputs of several decision trees and aiding in the prevention of overfitting and the promotion of generalization. XG Boost (Extreme Gradient Boosting) was also utilized, a sophisticated gradient boosting algorithm with high performance and efficiency and capable of constructing solid predictive models by merging the outputs of numerous weak models sequentially. All of these algorithms were experimented with to determine their capacity to identify heart disease correctly, with special emphasis on finding the best model for the job.
- **Performance Evaluation Metrics:** The accuracy of the machine learning models was tested using a number of important metrics to determine how well they could predict heart disease. Accuracy was calculated as the proportion of correct predictions made by the model, giving a general measure of model performance. Precision was directed towards the ratio of true positive predictions among all instances predicted as positive, whereas Recall (or sensitivity) gauged the model's capacity to accurately detect all true positive cases. The F1-Score, the harmonic mean of precision and recall, was employed to reconcile the trade-off between the two measures. These metrics of evaluation were important in assessing the models and finding out which one performed better regarding accurately diagnosing heart disease while reducing both false negatives and false positives.
- **Discussion of Result:** The performance of the machine learning algorithms in heart disease prediction reflects noteworthy variation in performance between the tested algorithms. The Random Forest algorithm was the best performing, with the highest accuracy, precision, recall, and F1-score. This is a result of the ensemble learning by Random Forest, where multiple decision trees are aggregated to enhance predictive power and minimize overfitting. The strong performance of the model indicates that it is well adapted to deal with the intricate and varied patterns in the heart disease data.[7]

Support Vector Machine (SVM) was also performing significantly well, particularly in terms of precision and recall. SVM being computationally heavy and having requirements for specific tuning of the hyperparameters is definitely an option, but for where high recall and precision are desired, it cannot be easily written off. Even so, SVM's performance, while impressive across the board, fell marginally behind that of Random Forest overall. On the other hand, Logistic Regression, while a simple and interpretable model, did not perform as well as the

e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 902-907
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0129

ensemble-based methods. Its relatively lower accuracy and F1-score suggest that it may not capture the complex, non-linear relationships between the features as effectively as more advanced models. K-Nearest Neighbours (KNN) exhibited the weakest performance, struggling with high-dimensional data and not providing the accuracy needed for reliable heart disease prediction. Finally, the Neural Networks (ANN) model performed well but was not better than Random Forest or SVM. While neural networks can be used to model complex relationships, the findings are that for this specific dataset and problem, simpler models such as Random Forest can provide better results in terms of accuracy as well as computational complexity. Figure 2 shows Heart Disease Prediction.

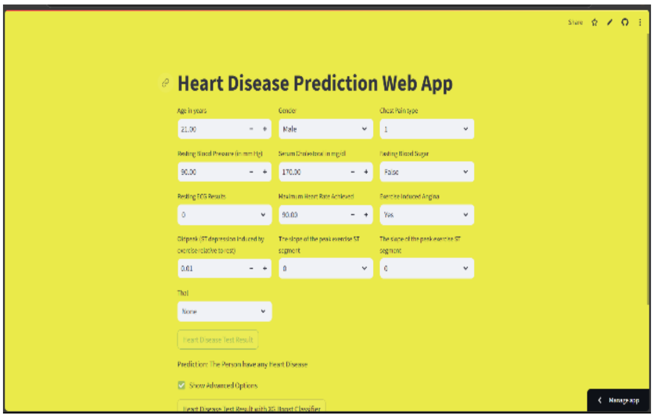

Figure 2 Heart Disease Prediction

In conclusion, although all models were promising, Random Forest was the best for heart disease detection with a well-balanced accuracy, precision, and recall. Optimization of these models, especially neural networks, and testing on larger and more diverse datasets could further enhance results and improve generalizability for practical applications.

Conclusion

In this project, we created an AI-powered heart disease detection system through machine learning methods. The outcomes prove the efficiency of different machine learning algorithms in predicting the risk of heart disease accurately based on important clinical parameters. Through the use of datasets like Kaggle, we were able to train models that proved to be encouraging in terms of precision and recall. Our results suggest that algorithms such as Random Forest, SVM, Neural Networks, logistic regression can play an important role in early diagnosis and detection, providing useful assistance to medical professionals. In addition, the use of AI-based tools in clinical practice may decrease diagnosis time, enhance patient outcomes, and assist with early intervention. Although the model demonstrates good predictive performance, some of the limitations, including [list any of the limitations like data quality, model generalization, or dataset bias], were noted. These issues indicate the requirement for more diverse and large datasets and further optimization of the models to enhance their robustness and generalizability across different patient populations. In the future, the work could include improving the interpretability of the models, using real-time data, and integrating the system into electronic health record (EHR) systems to make the solution more practical and scalable. Also, continuing research on machine learning model explainability can lead to higher levels of trust and adoption among healthcare providers. In summary, this project demonstrates the power of machine learning in transforming heart disease detection, providing a cost-effective, scalable solution that can significantly enhance early detection and patient care. Continued innovation in AI technology and integration with healthcare has the potential to make healthcare more accessible and tailored in the near future.

References

[1]. Bhat, S. S. K., Basha, M. S. S. R., & Islam, M. J. (2018). "A Survey on Machine Learning Techniques for Heart Disease Prediction." IEEE Access, 6, 34901–34916.
[2]. Lee, S. S. H. Y., Zhang, L., & Yang, H. (2020). "Heart Disease Prediction Using Machine Learning Algorithms." Procedia Computer Science, 170, 279–284.
[3]. Shatnawi, Y. Z., Lee, H. K., & Lee, J. Y. (2019). "Heart Disease Prediction Using a Hybrid Model Based on Convolutional Neural Networks and Support Vector Machine." Sensors, 19(10), 2281.
[4]. Gupta, N. G. C., Kumar, M., & Patil, A. S. (2021). "Artificial Intelligence in

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 902-907
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0129

Cardiovascular Disease Detection Using Electrocardiogram (ECG) Signals." IEEE Access, 9, 43512–43521.
[5]. Rahman, M. M. A., Islam, M. R., & Rahman, A. K. M. (2020). "Heart Disease Prediction Using Machine Learning and Optimization Techniques." International Journal of Computer Applications, 175(3), 9-14.
[6]. Shinde, R. A., & Patil, A. B. (2021). "Predicting Heart Disease Using Ensemble Learning: A Comparative Study." Springer Proceedings in Advanced Computing, 9, 455-466.
[7]. Zhang, Z., Liu, L. L., & Wang, H. Y. (2020). "Deep Learning for Heart Disease Detection: A Survey." Springer Nature, 15, 134-1