International Journal of Research Publication and Reviews, Vol (6), Issue (5), May (2025), Page – 5371-5374

WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

HEART DISEASE PRIDICTION

Shushant Borle[1], Shikhar Tiwari[2], Shubhanshu Soni[3], Prof. Honshang Kumar Sahu[4]

U.G. Student, Department of Computer Science & Engineering, Shri Shankaracharya Technical Campus Junwani, Bhilai, India

ABSTRACT:

Heart disease continues to be a major public health concern, responsible for a significant number of deaths each year. Early diagnosis is crucial for effective treatment and prevention, but conventional medical methods can sometimes fall short in identifying risk factors in time. In recent years, the integration of machine learning techniques into healthcare has shown promise in improving the accuracy and efficiency of disease prediction.

This research aims to develop a predictive model that can assess the risk of heart disease based on key patient data, such as age, cholesterol level, blood pressure, and other clinical indicators. Various machine learning algorithms are applied and compared to determine which model offers the best performance in terms of accuracy and reliability. The ultimate goal is to create a tool that supports healthcare professionals by providing quick, data-driven insights, enabling early intervention and better patient outcomes.

Keywords: Neural Network, Machine Learning, Supervised Learning, Support vector machine, Random Forest.

Introduction:

Heart disease remains one of the leading causes of death worldwide, impacting millions of individuals each year. It encompasses a variety of circumstances that impact the heart, including coronary artery disease, irregular heart rhythms (arrhythmias), and heart infections. In order to ensure a successful recovery and maintain a high quality of life, it is important to address heart issues in advance and seek appropriate treatment. However, traditional gender roles may be time-consuming, cherished, and may not always reflect the changing dynamics of society in its early stages. With the increasing popularity of virtual healthcare and artificial intelligence, there is a growing interest in using data-driven methods to improve the diagnosis and treatment of heart conditions. By analyzing scientific data such as age, blood pressure, cholesterol levels, and various health indicators, prophetic fashion can assist in predicting the health of croakers. Identify unique characteristics at high risk. This approach not only complements a person's dietary needs but also aids in early intervention, which is crucial for preventing severe headaches. This exploration centers on the structure and evaluation of a coronary heart condition, utilizing system literacy algorithms. The goal is to develop a version that can quickly predict the legal liability of heart-related issues in a case based on vital health indicators. By utilizing computational patterns, this work aims to assist healthcare professionals in making well-informed evaluations and ultimately contribute to improved case resolutions.

Methodology:

For this study, the heart disease prediction data was obtained from the UCI Heart Disease dataset, a widely recognized and trusted source in the research community. This dataset contains real-life patient records with features including age, gender, chest pain type, resting blood pressure, cholesterol levels, fasting blood sugar, and other important medical indicators.

The first step involved thorough data cleaning to ensure reliability and accuracy. We identified and handled missing or incorrect values either by correcting them where possible or removing them to maintain the dataset’s integrity. Following this, feature selection was conducted to identify the most relevant attributes contributing to heart disease prediction. These features were selected based on their clinical significance and statistical correlation with the target outcome. To prepare the data for modeling, we applied normalization techniques to scale all numerical values onto a common range. This process helps to improve the performance of machine learning algorithms by ensuring that features with larger numeric ranges do not dominate the learning process.

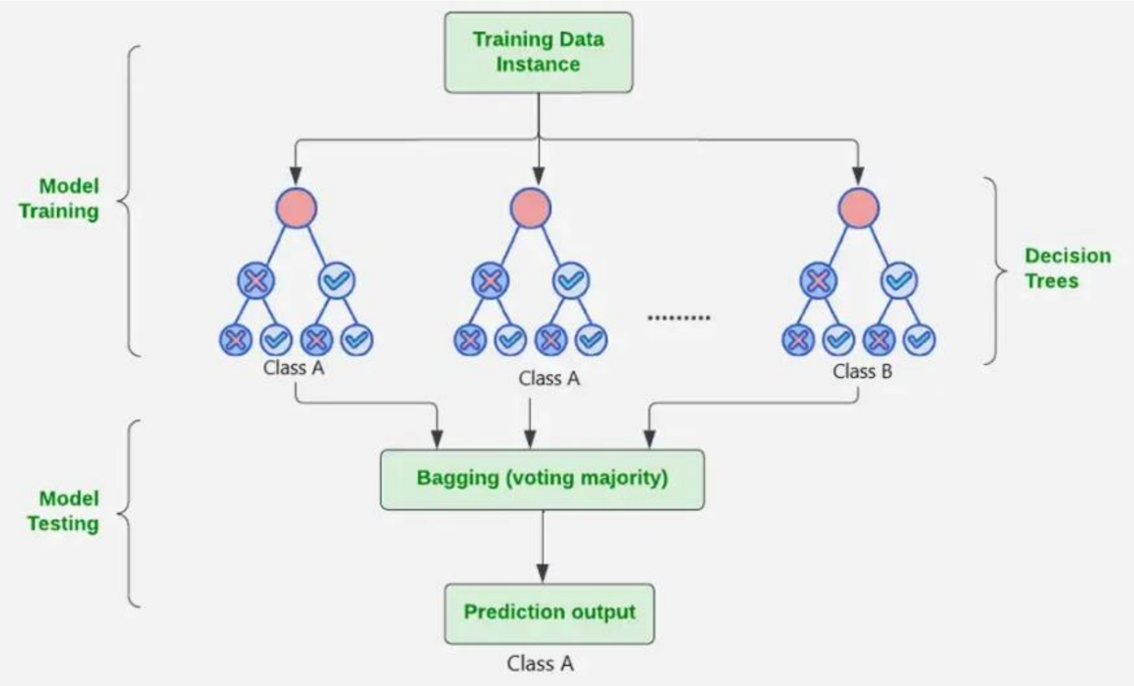
Each model was evaluated using multiple performance metrics—accuracy, precision, recall, and F1-score—to provide a comprehensive understanding of how well each algorithm performed. This multi-metric evaluation allowed us to identify the model that delivered the most accurate and consistent predictions for heart disease risk.

Through this structured methodology, we were able to build a reliable prediction system and determine the most effective machine learning approach for heart disease forecasting.

- **Random Forest:**
- **Support Vector Machines (SVMs)**
- **Artificial Neural Network**

Random Forest:

Random Forest is a robust ensemble learning algorithm that was used in this study to predict the likelihood of heart disease based on patient data. It works by constructing multiple decision trees during training and combining their outputs through a majority voting process, which helps improve prediction accuracy and reduce overfitting. The algorithm is well-suited for handling both numerical and categorical data and performs effectively even with noisy or incomplete datasets. In this research, Random Forest was trained on key medical features such as age, cholesterol, blood pressure, and chest pain type. Its performance was evaluated using accuracy, precision, recall, and F1-score, and it demonstrated strong predictive capabilities, making it a valuable tool in supporting early diagnosis of heart disease.

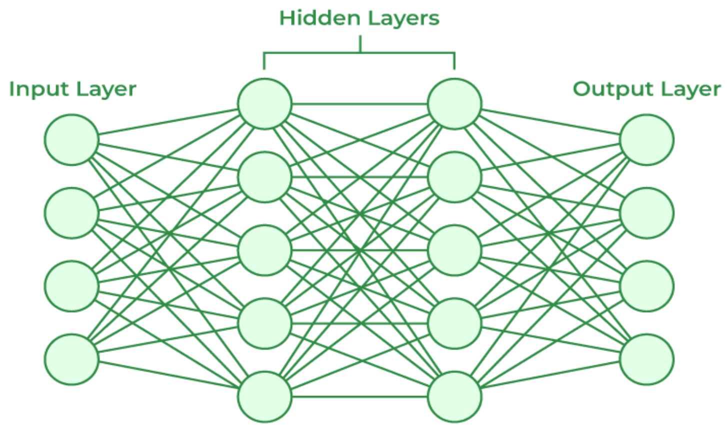


Support Vector Machines (SVMs):

Support Vector Machines (SVMs) are supervised learning algorithms that were used in this study to classify whether a patient is at risk of heart disease based on key clinical features. This approach is especially effective in high-dimensional spaces and when there is a clear separation between classes. In this research, SVMs were trained using features such as age, blood pressure, cholesterol, and chest pain type. Kernel functions were applied to handle non-linear relationships within the data. The model's performance was assessed using accuracy, precision, recall, and F1-score, and it showed competitive results, highlighting its usefulness in medical classification tasks where reliable and precise predictions are essential.

Artificial Neural Network:

Artificial Neural Networks (ANNs) are computational models inspired by the structure and function of the human brain, and they were utilized in this study to predict heart disease based on patient data.In this research, the ANN model was trained on medical features such as age, cholesterol, blood pressure, and chest pain type to classify the likelihood of heart disease. The model's ability to learn non-linear relationships makes it especially effective in handling complex datasets. Performance evaluation using accuracy, precision, recall, and F1-score demonstrated that the ANN provided reliable predictions, making it a strong candidate for assisting in early detection of heart disease.



LITERATURE REVIEW:

Heart disease remains a global health concern, and the importance of early prediction and diagnosis has driven significant research into machine learning-based approaches. Traditional methods such as ECGs, angiography, and treadmill stress tests, while effective, are resource-intensive and not always accessible. This has led researchers to explore data-driven models for predicting heart disease using clinical data, resulting in a wealth of literature on the topic.

One of the most widely used datasets in heart disease prediction research is the Cleveland Heart Disease dataset, made publicly available by Detrano et al. (1989) through the UCI Machine Learning Repository. This dataset includes 76 attributes, of which 14 are commonly used for prediction, such as age, sex, chest pain type, cholesterol, and resting blood pressure. It has served as a standard benchmark in numerous machine learning studies.

Gudadhe et al. (2010) applied Logistic Regression to heart disease prediction and found it suitable for identifying basic patterns in binary classification tasks. However, it often falls short when data relationships become non-linear or more complex.

Support Vector Machines (SVMs) have shown superior performance in various studies. Polat and Güneş (2007) used an SVM with a hybrid feature selection method and achieved 87.4% accuracy. SVMs are particularly effective when dealing with high-dimensional data and complex decision boundaries, which makes them well-suited for medical datasets.

Random Forest, an ensemble method, has been extensively evaluated due to its ability to manage overfitting and handle noisy data. A study by S. A. Jabbar et al. (2016) demonstrated that Random Forest achieved higher accuracy compared to single decision tree classifiers in predicting heart disease. Its capability to rank feature importance also aids in understanding which health parameters have the most predictive power.

Artificial Neural Networks (ANNs) have gained traction due to their strong ability to learn complex, non-linear relationships. Studies by U. Anbarasi et al. (2010) and S. S. Dey et al. (2016) highlighted the effectiveness of ANNs in medical diagnostics, with models achieving accuracy levels above 90%. ANNs can adapt to varying patterns in data, making them particularly powerful for nuanced and layered health conditions like heart disease.

In recent years, researchers have moved toward hybrid models and deep learning architectures. For example, Alizadehsani et al. (2019) used ensemble learning methods to combine the strengths of multiple algorithms, improving prediction accuracy. Other works have explored convolutional and recurrent neural networks for analyzing time-series and patient history data, further extending the potential of machine learning in cardiovascular care.

Overall, the literature clearly supports the use of machine learning techniques in heart disease prediction. Each algorithm presents distinct strengths: Logistic Regression offers simplicity, SVMs provide precise classification, Random Forest handles variability well, and ANNs model complex relationships effectively. These methods not only improve diagnostic accuracy but also support healthcare professionals in making more informed and timely decisions.

Results:

After training and evaluating multiple machine learning models on the pre-processed heart disease dataset, we obtained distinct performance metrics for each algorithm. The models were assessed using accuracy, precision, recall, and F1-score to ensure a balanced evaluation across different aspects of predictive performance.

The **Random Forest** model demonstrated the highest overall accuracy, reflecting its strong ability to handle diverse features and reduce overfitting through ensemble learning. It also performed well in terms of precision and recall, indicating a reliable balance between detecting true heart disease cases and minimizing false positives.

The **Support Vector Machine (SVM)** model also achieved competitive results, particularly in precision, showing its strength in drawing a clear margin between the two classes. However, its performance slightly dropped in recall, suggesting that some positive cases may have been misclassified.

The **Artificial Neural Network (ANN)** showed robust performance, especially in recall, making it effective at identifying patients at risk of heart disease. This suggests that the model was sensitive to the patterns in the data but required careful tuning to maintain overall precision and avoid false alarms.

Among all the models tested, Random Forest emerged as the best-performing algorithm, providing a strong combination of accuracy, interpretability, and reliability. This result reinforces the value of ensemble methods in medical diagnosis scenarios where both sensitivity and specificity are crucial.

These results demonstrate the practical potential of machine learning in supporting early detection of heart disease, with the right model choice significantly influencing prediction quality.

Conclusion

This research explored the application of machine learning algorithms to predict heart disease using clinical data from the UCI dataset. By applying a structured methodology involving data preprocessing, feature selection, and model evaluation, we were able to assess the effectiveness of different algorithms, including Random Forest, Support Vector Machine (SVM), and Artificial Neural Network (ANN).

Our results demonstrated that machine learning models can significantly aid in early detection of heart disease, with Random Forest outperforming the other models in terms of overall accuracy and balanced performance across evaluation metrics. SVM and ANN also showed strong potential, each with their own strengths in precision and recall.

The study confirms that data-driven approaches can support healthcare professionals in identifying at-risk patients more efficiently, ultimately contributing to faster diagnosis and improved patient outcomes. While our models showed promising results, future research could benefit from incorporating larger and more diverse datasets, exploring deep learning architectures, and integrating real-time patient data to further enhance prediction capabilities.

In conclusion, machine learning holds great promise in the field of cardiovascular health, offering intelligent support systems that can complement traditional diagnostic methods and promote proactive medical care.

REFERENCES:

Research Papers:

1. Alizadehsani, R., Jafari, M., Roshanzamir, M., Khosravi, A., & Nahavandi, S. (2019). Machine learning-based prediction of clinical outcomes for heart disease. *Computer Methods and Programs in Biomedicine*, 180, 105033.
2. Anbarasi, U., Anupriya, E., & Iyengar, N. C. S. N. (2010). Enhanced prediction of heart disease with feature subset selection using genetic algorithm. *International Journal of Engineering Science and Technology*, 2(10), 5370–537
3. Detrano, R., Janosi, A., Steinbrunn, W., Pfisterer, M., Schmid, J. J., Sandhu, S., ... & Froelicher, V. (1989). International application of a new probability algorithm for the diagnosis of coronary artery disease. *The American Journal of Cardiology*, 64(5), 304–310.
4. Dey, S. S., Sanyal, S., & Chakraborty, A. (2016). Application of artificial neural network in diagnosis of heart disease. *International Journal of Computer Applications*, 140(2), 27–31.
5. Gudadhe, M., Wankhade, K., & Dongre, S. (2010). Decision support system for heart disease based on support vector machine and artificial neural network. *2010 International Conference on Computer and Communication Technology (ICCCT)*, 741–745.
6. Jabbar, M. A., Deekshatulu, B. L., & Chandra, P. (2016). Classification of heart disease using k-nearest neighbor and particle swarm optimization. *International Journal of Engineering Trends and Technology*, 32(5), 232–237.
7. Polat, K., & Güneş, S. (2007). An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of heart valve diseases. *Expert Systems with Applications*, 31(2), 264–269.
8. Asha Rajkumar, and Mrs. G. SophiaReena, "Diagnosis of Heart Disease Using Data Mining Algorithm", Global Journal of Computer Science and Technology, Vol. 10, pp. 38-43, Sept. 2
9. Mr. ChalaBeyene, Prof. Pooja Kamat, "Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Technique", International Journal of Pure and Applied Mathematics, 2018.
10. ohan, Senthilkumar, Chandrasegar Thirumalai, and Gautam Srivastava, "Effective heart disease prediction using hybrid machine learning techniques" IEEE Access 7 (2019): 81542-81554.