International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Heart Disease Prediction using Machine Learning

[1]Sinchana M L(4MC22CS151) CSE & Malnad College Of Engineering (Hassan)
[2]Sinchana Raj G(4MC22CS153) CSE & Malnad College Of Engineering (Hassan)
[3]Mrunal C Shetty(4MC22CS101) CSE & Malnad College Of Engineering (Hassan)
[4]Venugopal B H(4MC22CS180) CSE & Malnad College Of Engineering (Hassan)
[5]Nayana R(Assistant Professor) CSE & Malnad College Of Engineering (Hassan)

Abstract - With the rise in cardiovascular diseases (CVDs), machine learning (ML) has become a key tool in early diagnosis using clinical data. Studies utilize algorithms like SVM, Random Forest, ANN, Decision Trees, and ensemble models, stressing the importance of data preprocessing and feature selection (e.g., PCA, RFE) for improved accuracy. Hybrid and ensemble techniques enhance predictive performance, while model interpretability remains vital for clinical use. Overall, ML models, when properly trained and tuned, significantly aid in the early detection of heart disease and support timely medical intervention.

Key Words: Heart Disease, Machine Learning, Feature Selection, Classification Algorithms, Data Prepro cessing, SMOTE, Explainable AI (SHAP), Accuracy Metrics (e.g., AUC, Precision, Recall)

1. INTRODUCTION

Heart disease is the leading global cause of death, influenced by lifestyle, genetics, and chronic issues. While traditional tools like ECGs are useful, they can be costly and less accessible. Machine learning (ML) offers a better alternative by analyzing complex clinical data for improved prediction. Models like Random Forest, SVM, XGBoost, and Neural Networks, along with feature selection methods, boost accuracy. Explainable AI, ensemble models, and mobile apps enable real-time, interpretable support, making diagnosis faster and more affordable.

2. LITERATURE SURVEY

The reviewed studies highlight the growing use of machine learning (ML) in heart disease prediction using various techniques and datasets. Bhatt et al. achieved strong accuracy with real-world data but lacked clinical depth. Biswas et al. compared feature selection methods effectively, though limited by a small sample. Hajiarbabi's meta-analysis emphasized methodology and diversity, while Ahire et al. offered a basic overview with minimal rigor. Common findings include the effectiveness of ensemble models, importance of feature selection, and challenges in generalizability. However, most studies lack clinical integration and interpretability. Future work must focus on validation, collaboration, and explainable AI for real-world impact.

1. **Machine Learning for Heart Disease Prediction**

   This body of work highlights the increasing use of machine learning (ML) models in predicting heart disease using clinical datasets. Techniques such as Random Forest, Support Vector Machines (SVM), and ensemble models were commonly employed. Bhatt et al. used a large real-world dataset and advanced preprocessing to achieve high accuracy, although the study lacked clinical interpretability. Biswas et al. provided a comparative analysis of feature selection techniques, showing that Random Forest performed well despite a limited sample size. While these models demonstrated strong technical results, most lacked clinical integration or real-world validation, reducing their practical impact.

2. **Meta-Analytical and Introductory Studies in Heart Disease Prediction**

   Hajiarbabi conducted a detailed meta-analysis focusing on methodology and dataset diversity, making it a strong academic reference. In contrast, Ahire et al. offered a basic, introductory overview of ML models for heart

disease, with minimal experimental evidence. Together, these studies emphasize the need for rigorous methodological approaches and deeper experimentation to support clinical applications. Despite the promising use of ML, the gap between research and clinical deployment remains a significant barrier.

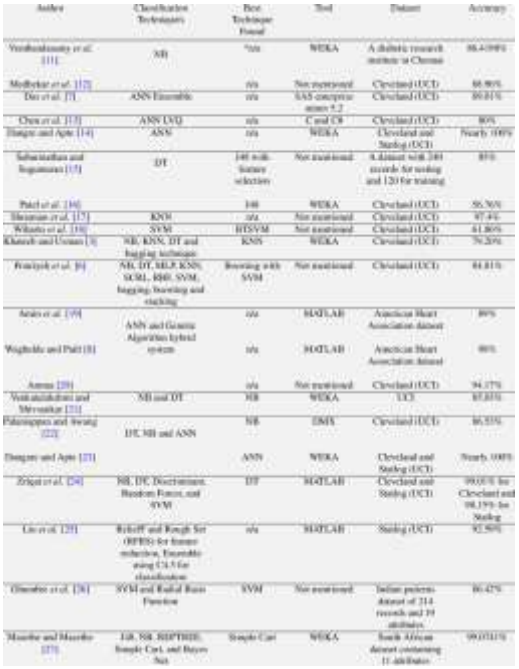
3. Common Themes and Future Directions

Across the reviewed literature, common themes emerged: the effectiveness of ensemble models, the critical role of feature selection, and the ongoing challenge of generalizability. However, few studies employed interpretability tools like SHAP or addressed real-world clinical integration. The lack of explainability and domain collaboration hinders trust and adoption in healthcare settings. Future research must focus on validation, explainable AI, and integration with clinical workflows to translate ML advancements into practical, patient-centered solutions

Table -1: Summary of Heart disease dataset

Dataset Name	Framingham Heart Disease Dataset		
Class	Two (1-Absent; 2-Presence)		
Instances	4012 (Absent-2945; Presence-1067)		
Attribute	10		
	Minimum Value	Maximum Value	Missing Values
Age	32	70	-
Gender (0-Male; 1-Female)	0	1	-
Blood Pressure (Systolic)	83.5	295	-
Blood Pressure (Diastolic)	48	142.5	-
Heart Rate	44	143	√
Blood Sugar	40	394	√
BMI	15.54	56.8	√
Cholesterol	107	696	√
Smoking (0-No; 1-Yes)	0	1	-

Table -2: Comparison of Classification techniques



3. OBJECTIVES

- **Evaluate ML Models for Heart Disease Prediction**

Assess different machine learning models like Random Forest, SVM, and Neural Networks based on accuracy, precision, recall, and AUC. This helps determine which model performs best across diverse datasets for reliable heart disease prediction.

- **Compare Feature Selection Techniques**

Use methods like PCA, RFE, ANOVA, and Chi-Square to identify the most relevant features. Effective feature selection reduces complexity and improves model performance and training efficiency.

- **Analyze Strengths and Limitations of ML Approaches**

Evaluate how well models generalize beyond training data, especially with limited or imbalanced datasets. Understand their practical limitations, such as lack of clinical context or overfitting in small samples.

- **Highlight Model Interpretability and Explainability**

Emphasize tools like SHAP and LIME to explain predictions in a way clinicians can understand. Interpretability builds trust and helps integrate AI into clinical decision-making.

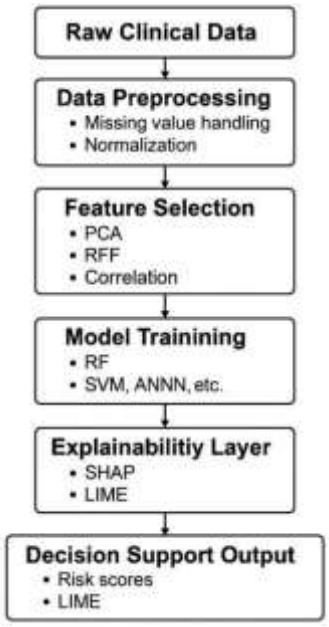
- **Identify Gaps in Clinical Integration**

Most ML models are not validated in real clinical settings or connected to hospital systems like EHRs. Bridging this gap is crucial for real-world usability and adoption in healthcare workflows.

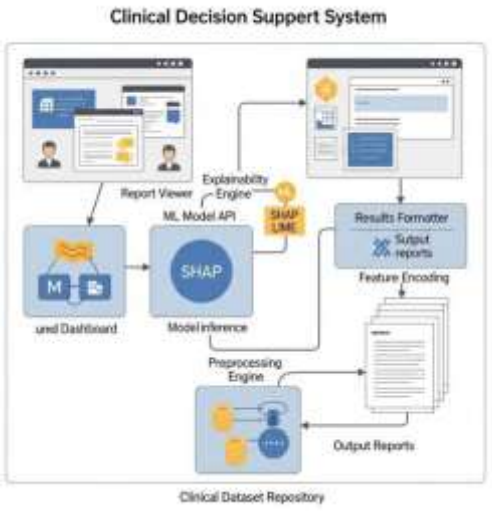
- **Provide Recommendations for Future Research**

Encourage use of larger datasets, cross-disciplinary collaboration, and real-time AI tools. Future studies should focus on making ML models clinically interpretable, scalable, and deployable.

Workflow diagram:



System Architecture:



4. EXPECTED OUTCOME

- **Accurate Heart Disease Prediction Model:**

  Develop a machine learning model that can reliably predict the presence or risk of heart disease with high accuracy, precision, recall, and AUC by comparing multiple algorithms like Random Forest, SVM, and Neural Networks.

- **Effective Feature Selection:**

  Identify the most relevant clinical and demographic features that contribute significantly to heart disease prediction by applying various feature selection techniques.

- **Model Interpretability and Explainability:**

  Incorporate explainability tools such as SHAP or LIME to ensure the model's predictions are transparent and understandable, helping clinicians trust and adopt the model for real-world applications.

- **Comprehensive Evaluation:**

  Evaluate the models on diverse datasets to validate robustness and generalizability, including analysis of strengths and limitations of the applied ML techniques in the context of heart disease diagnosis.

- **Clinical Relevance:**

  Provide insights into how the model's predictions can aid early diagnosis and risk assessment in clinical settings, potentially supporting better patient outcomes through timely intervention.

3. CONCLUSION

The papers highlight the growing role of machine learning in heart disease prediction. Bhatt et al. use a large dataset with strong accuracy but lack clinical depth. Biswas et al. apply feature selection with good results but a small dataset. The literature review offers comparative insights without new

experiments. Hajiarbabi provides a detailed academic review by data type but is less practical. Ahire et al. focus on basic ML education. Random Forest and MLP models perform well overall, but most studies lack clinical validation, explainability, and integration with healthcare professionals, limiting real-world use. Future work should focus on bridging this gap.

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to all the authors and researchers whose work served as the foundation for this project. Their significant contributions to the fields of 4 machine learning and heart disease prediction have provided critical insights that shaped our analysis. We are especially thankful to Bhatt et al., Biswas et al., Hajiarbabi, Ahire et al., and the contributors of the comparative literature review for their diverse methodologies and detailed research. We also acknowledge the value of open-access publishing, which made this comparative study possible. Furthermore, we extend our heartfelt thanks to our mentors, peers, and academic community for their guidance, support, and constructive feedback throughout the development of this project.

REFERENCES

[1] Baldonado, M., Chang, C.-C.K., Gravano, L., Paepcke, A.: The Stanford Digital Library Metadata Architecture. Int. J. Digit. Libr. 1 (1997) 108–121

[2] Bruce, K.B., Cardelli, L., Pierce, B.C.: Comparing Object Encodings. In: Abadi, M., Ito, T. (eds.): Theoretical Aspects of Computer Software. Lecture Notes in Computer Science, Vol. 1281. Springer-Verlag, Berlin Heidelberg New York (1997) 415–438

[3] van Leeuwen, J. (ed.): Computer Science Today. Recent Trends and Developments. Lecture Notes in Computer Science, Vol. 1000. Springer-Verlag, Berlin Heidelberg New York (1995)