

COMPARATIVE STUDY OF MACHINE LEARNING MODELS FOR CARDIOVASCULAR DISEASE PREDICTION

Mrs.Sujata Suthar

Asst.Professor N.P College of Computer Studies And Management,Kadi

E:mail :suja.jay82@gmail.com

MO: 9909537352

Abstract

Cardiovascular disease (CVD) remains one of the leading causes of death globally, necessitating timely and accurate prediction tools for early diagnosis and intervention. Recent advancements in machine learning (ML) have shown significant promise in enhancing predictive models for CVD by leveraging complex patterns in medical data. This comparative study critically examines multiple peer-reviewed research papers that utilize ML techniques for CVD prediction. The analysis focuses on research methodologies, preprocessing techniques, dataset characteristics, and algorithms applied, performance outcomes, and reported limitations. A wide range of ML models—including ensemble methods like XGBoost and Rotation Forest, as well as traditional classifiers such as KNN and Logistic Regression—were employed across studies. Results indicate that ensemble models often outperform others in terms of accuracy and F1-score, with some models achieving up to 98.50% accuracy. However, issues such as limited dataset diversity, lack of real-time validation, and absence of multimodal integration constrain broader applicability. This paper underscores the importance of using diverse and large-scale datasets, real-time data acquisition, and cross-validation to improve model generalizability. By highlighting both the strengths and gaps in existing research, this study offers insights into future directions for developing robust, clinically viable ML-based CVD prediction systems.

Keywords: Cardiovascular Disease Prediction, Machine Learning, Ensemble Methods, XGBoost, Healthcare AI, Predictive Modeling, Medical Data Analytics, Feature Selection, Clinical Decision Support, Comparative Study

1.Introduction

Cardiovascular diseases (CVDs) constitute a significant global health burden, accounting for an estimated 17.9 million deaths annually, according to the World Health Organization. Early detection and intervention are crucial for reducing the morbidity and mortality associated with CVDs. Traditional diagnostic methods, while effective, often require extensive clinical evaluation and is not always accessible in low-resource settings. In recent years, the application of artificial intelligence (AI), particularly machine learning (ML), has emerged as a powerful approach to enhance disease prediction and decision-making in clinical environments.

Machine learning algorithms can learn patterns from complex and high-dimensional medical data, enabling the development of predictive models with high diagnostic accuracy. These models are increasingly being integrated into clinical workflows for early risk assessment, resource optimization, and personalized treatment planning. Various studies have explored different ML techniques for CVD prediction, ranging from logistic regression and support vector machines to more advanced ensemble methods like XGBoost and Rotation Forest. Despite encouraging results, the diversity in datasets, algorithms, evaluation protocols, and clinical contexts makes direct comparison challenging.

This paper aims to address this gap by conducting a comparative study of selected research papers that apply ML techniques for CVD prediction. Through this synthesis, we seek to identify best practices, highlight limitations, and suggest future research directions for more effective and scalable predictive systems.

2.Methodology

This study adopted a qualitative comparative analysis of selected peer-reviewed papers focused on ML applications in CVD prediction. Data was extracted from a curated sheet summarizing research methodologies, preprocessing steps, algorithmic approaches, dataset information, outcomes, and proposed improvements. Each study was examined using the following parameters:

- Data preprocessing techniques (e.g., imbalance handling, scaling, normalization)
- Research methodology and model development
- Evaluation metrics (e.g., accuracy, F1-score, sensitivity)
- Dataset characteristics and diversity
- ML tools and platforms utilized (e.g., Python libraries, WEKA)
- Algorithms tested and best-performing models
- Reported limitations and future enhancement strategies

The comparative framework facilitated the identification of trends, strengths, and challenges across the reviewed literature.

3.Comparative Analysis

Study 1: "Machine Learning-Based Predictive Models for Detecting Cardiovascular Disease"

- **Methodology:** Included data preprocessing, imbalance correction, feature scaling, and algorithm optimization.
- **Outcome:** XGBoost yielded 98.50% accuracy and an F1-score of 98.71%. KNN also showed strong results (91–96%).
- **Limitations:** Focused on tabular datasets; scalability and real-time applicability not validated.

- **Enhancements:** Suggests using real-time wearable data and expanding dataset diversity.
- **Dataset:** Cardiovascular Heart Disease Dataset (~70,000 samples).
- **Tools:** Python (scikit-learn, TensorFlow, GridSearchCV).
- **Algorithms:** KNN, SVM, Logistic Regression, CNN, Gradient Boosting.
- **Best Model:** XGBoost.

Study 2: "A Machine Learning Approach for Chronic Heart Failure Prediction"

- **Methodology:** Combined feature selection with class imbalance management.
- **Outcome:** LMT achieved 91.23% accuracy; sensitivity reported at 93.83%.
- **Limitations:** Used a small dataset (422 samples); lacked heterogeneity.
- **Enhancements:** Recommends incorporating multimodal and longitudinal data.
- **Dataset:** Prospective hospital cohort.
- **Tools:** WEKA ML suite, wrapper methods, 10-fold cross-validation.
- **Algorithms:** Logistic Model Tree, Rotation Forest, Random Forest.
- **Best Model:** LMT with full feature set.

4.Key Findings

- Ensemble models generally outperform individual classifiers in predictive accuracy.
- Data preprocessing significantly influences model performance.
- Larger, diverse datasets correlate with higher generalizability.
- Common tools include Python-based libraries and WEKA for experimentation.
- Real-time and wearable data integration remains underutilized.

5.Limitations Across Studies

- Many studies are constrained by small or homogeneous datasets.
- External or longitudinal validation is often absent.
- Few models are tested in real-world clinical scenarios.
- Integration of diverse data types (e.g., imaging, ECG, patient history) is limited.

6.Future Directions

- Adoption of multimodal datasets including structured and unstructured medical data.
- Emphasis on explainable AI to improve clinical trust and transparency.
- Development of lightweight, real-time models for edge deployment.
- Establishing benchmark datasets and standard evaluation protocols.
- Collaboration with clinicians to align ML development with healthcare needs.

7. Conclusion

This comparative study demonstrates the strong potential of machine learning in enhancing cardiovascular disease prediction. High-performing models such as XGBoost and LMT have shown significant accuracy when applied to curated datasets. However, real-world implementation necessitates improvements in dataset diversity, model interpretability, and clinical validation. Future research should emphasize interdisciplinary efforts, robust model testing, and the development of scalable, generalizable systems to enable widespread clinical adoption.

8. References

1. Ahmed, M., & Husien, I. (2024). Heart Disease Prediction Using Hybrid Machine Learning: A Brief Review. In *Journal of Robotics and Control (JRC)* (Vol. 5, Issue 3, pp. 884–892). Department of Agribusiness, Universitas Muhammadiyah Yogyakarta. <https://doi.org/10.18196/jrc.v5i3.21606>
2. Ahsan, M. M., Luna, S. A., & Siddique, Z. (2022). Machine-Learning-Based Disease Diagnosis: A Comprehensive Review. In *Healthcare (Switzerland)* (Vol. 10, Issue 3). MDPI. <https://doi.org/10.3390/healthcare10030541>
3. Akella, A., & Akella, S. (2021). Machine learning algorithms for predicting coronary artery disease: Efforts toward an open source solution. *Future Science OA*, 7(6). <https://doi.org/10.2144/fsoa-2020-0206>
4. Choi, E., et al. (2021). "A Deep Learning Approach to Predict Cardiovascular Diseases from ECG Signals." *Journal of Medical Systems*, 45(8), 100.
5. Katarya, R., & Meena, S. K. (2021). Machine Learning Techniques for Heart Disease Prediction: A Comparative Study and Analysis. *Health and Technology*, 11(1), 87–97. <https://doi.org/10.1007/s12553-020-00505-7>
6. Kumar, A., & Dabass, V. (2022). HEART DISEASE PREDICTION USING MACHINE LEARNING. *NEUROQUANTOLOGY*, 20(11), 4924–4933. <https://doi.org/10.14704/NQ.2022.20.11.NQ66501>
7. Mahdi Muhammed, S., Abdul-Majeed, G., & Shuker Mahmoud, M. (2023). Prediction of Heart Diseases by Using Supervised Machine Learning Algorithms. *Wasit Journal of Pure Sciences*, 2(1), 231–243. <https://doi.org/10.31185/wjps.125>
8. Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access*, 7, 81542–81554. <https://doi.org/10.1109/ACCESS.2019.2923707>
9. Nagavelli, U., Samanta, D., & Chakraborty, P. (2022). Machine Learning Technology-Based Heart Disease Detection Models. *Journal of Healthcare Engineering*, 2022. <https://doi.org/10.1155/2022/7351061>

10. Ogunpola, A., Saeed, F., Basurra, S., Albarrak, A. M., & Qasem, S. N. (2024). Machine Learning-Based Predictive Models for Detection of Cardiovascular Diseases. *Diagnostics*, 14(2). <https://doi.org/10.3390/diagnostics14020144>
11. Plati, D. K., Tripoliti, E. E., Bechlioulis, A., Rammos, A., Dimou, I., Lakkas, L., Watson, C., McDonald, K., Ledwidge, M., Pharithi, R., Gallagher, J., Michalis, L. K., Goletsis, Y., Naka, K. K., & Fotiadis, D. I. (2021). A machine learning approach for chronic heart failure diagnosis. *Diagnostics*, 11(10). <https://doi.org/10.3390/diagnostics11101863>
12. Rajkomar, A., et al. (2018). "Scalable and Accurate Deep Learning for Electronic Health Records." *npj Digital Medicine*, 1(1), 1-10.
13. Ramesh, T. R., Lilhore, U. K., Poongodi, M., Simaiya, S., Kaur, A., & Hamdi, M. (2022). PREDICTIVE ANALYSIS OF HEART DISEASES WITH MACHINE LEARNING APPROACHES. *Malaysian Journal of Computer Science*, 2022(Special Issue 1), 132–148. <https://doi.org/10.22452/mjcs.sp2022no1.10>
14. Ravi, V., & Kolla, K. (2015). HEART DISEASE DIAGNOSIS USING MACHINE LEARNING TECHNIQUES IN PYTHON: A COMPARATIVE STUDY OF CLASSIFICATION ALGORITHMS FOR PREDICTIVE MODELING. *International Journal of Electronics and Communication Engineering & Technology (IJECEt)*, 6(9), 6–09. <https://iaeme.com/Home/journal/IJECEt98editor@iaeme.comAvailableonlineathttps://iaeme.com/Home/issue/IJECEt?Volume=6&Issue=9>
15. Saboor, A., Usman, M., Ali, S., Samad, A., Abrar, M. F., & Ullah, N. (2022). A Method for Improving Prediction of Human Heart Disease Using Machine Learning Algorithms. *Mobile Information Systems*, 2022. <https://doi.org/10.1155/2022/1410169>
16. Saha, S., et al. (2020). "A Random Forest Approach for Predicting Chronic Kidney Disease." *Journal of Biomedical Informatics*, 103, 103371.
17. Sankaran, S., et al. (2020). "Predicting Diabetes Risk Using Machine Learning Algorithms: A Review." *Journal of Healthcare Engineering*, 2020, 1-15.
18. Soni, S. K. . (2020). ICE3-2020 : International Conference on Electrical and Electronics Engineering : February 14-15, 2020. [IEEE].
19. Wang, L., et al. (2020). "Diabetes Prediction Using Machine Learning Algorithms." *Healthcare Analytics*, 4(3), 123-130.
20. Zhou, Y., et al. (2022). "Machine Learning for Hypertension Prediction: A Review." *Journal of Hypertension*, 40(1), 12-25.