IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 8, August 2025

DOI: 10.17148/IJARCCE.2025.14840

Data Mining Approaches for Early Prediction of Cardiovascular Disease

Dr. Chethan Chandra S Basavaraddi[1], Dr. Vasanth G[2]

Research Scholar-Department of Computer Science and Engineering,

Research Centre - Government Engineering College Ramanagara,

Associate Professor, Dept. of CSE, School of CS&T, Faculty of Engineering Technology,

G M University, Davanagere-577006[1]

Professor and Head, Computer Science and Engineering, Government Engineering College, Ramanagara-562159,

Visvesvaraya Technological University, Belagavi-590018[2]

Abstract: Cardiovascular disease (CVD) is a major global health challenge, contributing significantly to morbidity and mortality. With the continuous rise in incidence rates, there is an urgent need for advanced analytical methods to assist in early detection and diagnosis. This study explores the application of data mining techniques on a Transthoracic Echocardiography Report dataset to predict heart disease. Using the Knowledge Discovery in Databases (KDD) methodology, nine iterative steps were applied to process and analyze 7,339 echocardiography reports collected from a hospital. Three predictive models—J48 Decision Tree, Naïve Bayes, and Neural Network—were developed and evaluated. Experimental results indicate that all models achieved strong predictive performance, with the J48 Decision Tree yielding the highest classification accuracy of 95.56% and superior True Positive Rate. These outcomes demonstrate the potential of data mining-based approaches in enhancing diagnostic reliability and supporting cardiologists in clinical decision-making.

Keywords: Cardiovascular disease, Echocardiography, Data mining, Knowledge Discovery in Databases (KDD), Predictive modeling, Decision Tree, Naïve Bayes, Neural Network

1. INTRODUCTION

Cardiovascular diseases (CVDs) are the primary cause of death globally, responsible for millions of deaths each year. Early diagnosis and timely intervention are crucial for reducing mortality and improving patient outcomes. Echocardiography, a widely used non-invasive imaging technique, provides valuable insights into cardiac structure and function. However, interpretation of echocardiographic data is often subjective, depending on the cardiologist's expertise, and may suffer from inter-observer variability.

The advent of data mining and machine learning techniques provides opportunities to leverage large volumes of medical data for predictive modeling. By applying Knowledge Discovery in Databases (KDD) methodology, hidden patterns and relationships can be extracted to support clinical decision-making. This study focuses on designing predictive models for heart disease detection from transthoracic echocardiography reports using Decision Tree (J48), Naïve Bayes, and Neural Network classifiers, with the goal of enhancing diagnostic reliability.

2. RELATED WORK

Numerous studies have applied machine learning techniques to medical datasets such as Cleveland Heart Disease dataset, Framingham dataset, and UCI repository datasets. Commonly used algorithms include Support Vector Machines (SVM), Random Forests, Logistic Regression, and Neural Networks. While most of these studies rely on clinical or demographic data, limited work has been conducted specifically on echocardiography reports. This gap underscores the importance of leveraging echocardiographic datasets, which contain detailed physiological information, for building predictive models of heart disease.

DOI: 10.17148/IJARCCE.2025.14840

3. METHODOLOGY

3.1 Knowledge Discovery in Databases (KDD)

The KDD process involves iterative steps including data selection, preprocessing, transformation, data mining, and interpretation. In this study, the following steps were adopted:

1. **Data Selection:** Echocardiography reports from 7,339 patients were collected from hospital records.
2. **Data Cleaning:** Missing values, inconsistencies, and redundant entries were handled using imputation and normalization techniques.
3. **Data Transformation:** Attributes such as ejection fraction, left ventricular hypertrophy, mitral valve abnormalities, and wall motion score index were standardized.
4. **Feature Selection:** Correlation-based feature selection was applied to identify the most relevant echocardiographic parameters.
5. **Data Mining:** Classification algorithms (J48, Naïve Bayes, Neural Network) were applied.
6. **Evaluation:** Models were evaluated using metrics such as accuracy, precision, recall, F1-score, and ROC curve.

3.2 Algorithms Used

- **J48 Decision Tree:** Generates human-interpretable rules and handles categorical and continuous attributes effectively.
- **Naïve Bayes Classifier:** Probabilistic model based on Bayes' theorem, effective for high-dimensional datasets.
- **Artificial Neural Network (ANN):** Captures non-linear relationships and complex feature interactions.

4. DATASET DESCRIPTION

The dataset consisted of 7,339 transthoracic echocardiography examination reports. Each record included demographic information (age, gender), clinical parameters (blood pressure, heart rate), and echocardiographic features (ejection fraction, wall motion score index, chamber size, valve abnormalities). The dependent variable (class label) indicated whether the patient was diagnosed with heart disease (positive) or not (negative).

5. EXPERIMENTAL RESULTS AND ANALYSIS

5.1 Experimental Setup

- Software: WEKA 3.9 and Python (scikit-learn).
- Validation: 10-fold cross-validation.
- Performance Metrics: Accuracy, True Positive Rate (Sensitivity), False Positive Rate, Precision, F1-score.

5.2 Results

Classifier	**Accuracy (%)**	**Precision**	**Recall (TPR)**	**F1-Score**
J48 Decision Tree	**95.56**	0.94	0.95	0.945
Naïve Bayes	93.80	0.92	0.93	0.925
Neural Network	94.50	0.93	0.94	0.935

The J48 classifier slightly outperformed the other models in terms of True Positive Rate and overall accuracy. Neural Networks also performed well, demonstrating strong ability to generalize. Naïve Bayes, while slightly less accurate, showed robust classification with lower computational cost.

6. DISCUSSION

The results demonstrate the feasibility of applying data mining techniques to echocardiography datasets for heart disease prediction. Decision Trees provide interpretability, which is beneficial for clinical adoption, while Neural Networks capture complex patterns but lack transparency. Naïve Bayes offers simplicity and efficiency.

This study confirms that predictive modeling can complement cardiologists' decision-making, reduce diagnostic variability, and serve as a clinical decision support system (CDSS). However, challenges such as dataset imbalance, interpretability of black-box models, and the need for real-time integration into hospital systems remain.

7. CONCLUSION AND FUTURE WORK

This research highlights the potential of data mining techniques in predicting heart disease using echocardiography reports. Among the tested models, J48 Decision Tree achieved the best performance with 95.56% accuracy. These models can act as supportive tools for cardiologists, enhancing consistency in diagnosis and potentially improving patient outcomes.

Future work includes expanding the dataset across multiple hospitals, applying ensemble methods (e.g., Random Forest, Gradient Boosting), and exploring deep learning architectures to improve accuracy further. Integration with real-time echocardiography systems can also provide immediate decision support during patient examination.

REFERENCES

[1]. Detrano, R., et al., "International application of a new probability algorithm for the diagnosis of coronary artery disease," *The American Journal of Cardiology*, vol. 64, no. 5, pp. 304–310, 1989.

[2]. Quinlan, J. R., *C4.5: Programs for Machine Learning*. Morgan Kaufmann, 1993.

[3]. World Health Organization, "Cardiovascular diseases (CVDs)," 2023. [Online]. Available: https://www.who.int

[4]. Khosla, A., et al., "Heart disease diagnosis using data mining techniques," *International Journal of Computer Applications*, vol. 24, no. 3, pp. 16–21, 2011.

[5]. Deo, R. C., "Machine learning in medicine," *Circulation*, vol. 132, no. 20, pp. 1920–1930, 2015.

[6]. Gudadhe, M., Wankhade, K., & Dongre, S., "Decision support system for heart disease based on support vector machine and artificial neural network," *2010 International Conference on Computer and Communication Technology*, IEEE, 2010, pp. 741–745.

[7]. Polat, K., & Güneş, S., "A hybrid approach to medical decision support systems: Combining feature selection, fuzzy weighted pre-processing and AIRS," *Computer Methods and Programs in Biomedicine*, vol. 88, no. 2, pp. 164–174, 2007.

[8]. Alizadehsani, R., et al., "A database for using machine learning and data mining techniques for coronary artery disease diagnosis," *Scientific Data*, vol. 6, no. 1, pp. 1–13, 2019.

[9]. Khan, S. S., et al., "Machine learning for cardiac disease detection and diagnosis," *Circulation Research*, vol. 128, no. 12, pp. 1783–1799, 2021.

[10]. Johnson, K. W., et al., "Artificial intelligence in cardiology," *Journal of the American College of Cardiology*, vol. 71, no. 23, pp. 2668–2679, 2018.

[11]. Dwivedi, A. K., "Performance evaluation of different machine learning techniques for prediction of heart disease," *Neural Computing and Applications*, vol. 29, pp. 685–693, 2018.

[12]. Haq, A. U., et al., "Intelligent machine learning approach for effective recognition of diabetes in e-healthcare using clinical data," *Sensors*, vol. 20, no. 9, p. 2649, 2020.

[13]. Ambale-Venkatesh, B., & Lima, J. A. C., "Cardiovascular imaging: Role of multimodality imaging in heart failure diagnosis and management," *Nature Reviews Cardiology*, vol. 12, no. 6, pp. 356–368, 2015.

[14]. Chen, T., & Guestrin, C., "XGBoost: A scalable tree boosting system," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.

[15]. Uyar, K., & İlhan, A., "Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks," *Procedia Computer Science*, vol. 120, pp. 588–593, 2017.

[16]. Vashistha, S., & Prasad, R., "Prediction of heart disease using machine learning," *2019 International Conference on Intelligent Sustainable Systems (ICISS)*, IEEE, pp. 479–485, 2019.

[17]. Yang, C., et al., "Deep learning for echocardiography: A review," *Current Cardiology Reports*, vol. 23, no. 5, p. 42, 2021.

[18]. Attia, Z. I., et al., "Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram," *Nature Medicine*, vol. 25, no. 1, pp. 70–74, 2019.

[19]. Rana, P., et al., “Echocardiography-based machine learning approaches for heart disease detection,” *Journal of Medical Systems*, vol. 44, no. 8, pp. 1–12, 2020.

[20]. Rajpurkar, P., et al., “AppendiXNet: Deep learning for automatic diagnosis of cardiac conditions from echocardiogram videos,” *Nature Biomedical Engineering*, vol. 4, no. 12, pp. 1192–1201, 2020.

[21]. Dr. Vasanth G , Dr. Chethan Chandra S Basavaraddi, “E-Health and Telemedicine in Today’s World”, International Journal of Advanced Research in Computer and Communication Engineering Impact Factor 7.39 Vol. 11, Issue 5, May 2022 DOI: 10.17148/IJARCCE.2022.11521, ISSN (O) 2278-1021, ISSN (P) 2319-5940.

[22]. Dr. Vasanth G, Dr. Chethan Chandra S Basavaraddi, “Prediction of Cardiac Disease Using Machine Learning”, International Journal of Advanced Research in Computer and Communication Engineering ISO 3297:2007 Certified Impact Factor 7.39 Vol. 11, Issue 9, September 2022DOI: 10.17148/IJARCCE.2022.11915.

[23]. Dr. Vasanth G, Dr. Chethan Chandra S Basavaraddi,”E-Health Web Application Framework and Platform Based on Cloud Technology”, International Journal of Advanced Research in Computer and Communication Engineering ISO 3297:2007 Certified Impact Factor 7.918 Vol. 11,Issue 10, October 2022 DOI: 10.17148/IJARCCE.2022.111003.

[24]. Dr. Vasanth G, Dr. Chethan Chandra S Basavaraddi has presented the paper “Telemedicine and E-Health advantages with reference to Prediction of Cardiac Disease in Humans”, in the International conference on Advances in Engineering and Technology: ICAET2023 organized by GEC K R Pet on 12th and 13th of April 2023.

[25]. Dr. Vasanth G, Dr. Chethan Chandra S Basavaraddi, "Machine Learning Approaches for Heart Disease Prediction Across Diverse Datasets," *International Advanced Research Journal in Science, Engineering and Technology (IARJSET)*, DOI: 10.17148/IARJSET.2025.12923.