International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 10 No. 1, 2024
https://theaspd.com/index.php

An Explainable Hybrid WCBA–BPN–Fuzzy Framework for Heart Disease Prediction: A Comparative Performance Analysis

Anuradha Diwan[1], Sanjeev Karmakar[2] , Sunita Soni[3]
[1]Department of Computer Applications, Bhilai Institute of Technology, Durg, Chhattisgarh, India
[2]Department of Computer Applications, Bhilai Institute of Technology, Durg, Chhattisgarh, India
[3]Department of Computer Science and Engineering, Bhilai Institute of Technology, Durg, Chhattisgarh, India
Email: [1]anuradha.shinestar@gmail.com, [2]dr.karmakars@gmail.com, [3]sunitasoni74@gmail.com

ABSTRACT
Heart disease remains one of the leading causes of mortality worldwide, necessitating the development of accurate and interpretable prediction systems for early diagnosis. Traditional machine learning approaches often achieve satisfactory predictive performance but lack transparency and the ability to effectively manage uncertainty present in clinical data. To address these limitations, this study proposes an explainable hybrid framework that integrates Weighted Classification Based Association (WCBA), Back Propagation Neural Network (BPNN), and Fuzzy Logic for heart disease prediction. The Cleveland Heart Disease dataset obtained from the UCI Machine Learning Repository was used for experimental evaluation. The proposed framework employs weighted association rule mining to identify clinically significant features, a neural network to model complex nonlinear relationships among risk factors, and fuzzy reasoning to handle uncertainty and provide interpretable diagnostic outcomes. A weighted fusion mechanism combines the outputs of the three modules to generate the final prediction. Experimental results obtained through stratified 10-fold cross-validation demonstrate that the proposed framework achieves superior performance compared with conventional machine learning models. The model attained an accuracy of 92.08%, precision of 0.91, recall of 0.93, F1-score of 0.92, and ROC-AUC of 0.95. The integration of feature-weighted learning and fuzzy inference significantly improves both predictive capability and interpretability. The proposed framework can serve as a reliable clinical decision-support system for the early detection of cardiovascular diseases and may be extended to other healthcare prediction applications.
Keywords*: Heart Disease Prediction, Explainable AI, WCBA, Neural Network, Fuzzy Logic, Clinical Decision Support System.*

1. INTRODUCTION

Cardiovascular diseases remain among the most significant causes of mortality worldwide. According to global health reports, millions of individuals suffer from coronary artery disease, myocardial infarction, and other cardiac disorders each year. Early identification of high-risk patients enables timely intervention and significantly improves survival rates.

Recent advances in artificial intelligence and machine learning have transformed healthcare analytics. Techniques such as Logistic Regression, Support Vector Machines, Random Forests, and Artificial Neural Networks have been extensively used for disease prediction. Although these methods demonstrate promising predictive capability, challenges related to interpretability, uncertainty management, and feature relevance continue to limit their practical adoption in clinical environments. To overcome these limitations, this study introduces an integrated hybrid framework combining Weighted Classification Based Association, Back Propagation Neural Networks, and Fuzzy Logic. The framework leverages the strengths of rule-based reasoning, nonlinear learning, and uncertainty handling to achieve accurate and interpretable heart disease prediction.

Major contributions include:

- Development of a hybrid explainable prediction framework.
- Integration of weighted association rule mining and neural learning.
- Incorporation of fuzzy reasoning for uncertainty management.
- Comprehensive performance evaluation using standard metrics.
- Improved interpretability for clinical decision support.

2. LITERATURE REVIEW

The rapid advancement of intelligent healthcare technologies has substantially improved the early detection and diagnosis of cardiovascular diseases, which continue to represent a major cause of mortality worldwide. In recent years, machine learning and soft computing approaches, including Back-Propagation Neural Networks (BPNN), Classification Based on Association (CBA), and Fuzzy Logic Systems, have been extensively employed for heart disease prediction. Although these techniques have demonstrated promising predictive capabilities, individual models frequently encounter challenges such as overfitting, limited interpretability, rule redundancy, and inadequate management of uncertainty inherent in clinical datasets. Consequently, researchers have increasingly focused on developing hybrid frameworks that combine the complementary strengths of multiple computational techniques.

BPNN models are particularly effective in learning complex non-linear relationships and optimising network parameters through adaptive weight adjustment. In contrast, Weighted Classification Based on Association (WCBA) enhances classification performance by extracting and prioritising meaningful association rules. Fuzzy Logic contributes an additional layer of interpretability by modelling uncertainty and imprecision commonly associated with clinical decision-making. The integration of these methodologies has the potential to improve predictive accuracy, robustness, and clinical reliability. This section critically examines previous studies related to hybrid intelligent systems for heart disease prediction and establishes the foundation for the proposed framework.

Bhuiyan et al. (2025) [15] conducted a comparative study involving Fuzzy Logic, Neural Networks, and Case-Based Reasoning (CBR) techniques for heart disease prediction using the UCI Heart Disease dataset. Their findings demonstrated that intelligent decision-support systems could achieve high predictive performance, with the best-performing model attaining an accuracy of 91.9%.

Diwan et al. (2025) [16] proposed a hybrid framework combining Back-Propagation Neural Networks and Classification Based on Association for heart disease diagnosis. The study reported a prediction accuracy of 89.44%, outperforming standalone BPNN and associative classification approaches. Furthermore, the hybrid model exhibited improvements in precision, recall, and F1-score, indicating enhanced predictive effectiveness.

Kaur and Khehra (2021) [17] developed a neuro-fuzzy framework for heart attack prediction by integrating neural network learning with fuzzy inference mechanisms. Their work highlighted the advantages of combining adaptive learning with interpretable reasoning, providing a foundation for more advanced hybrid architectures.

Aghamohammadi et al. (2019) [18] introduced a hybrid prediction system incorporating Genetic Algorithms (GA), Adaptive Neuro-Fuzzy Inference Systems (ANFIS), and K-fold cross-validation. Using the Cleveland Heart Disease dataset, the authors evaluated model performance across fourteen clinical attributes and achieved accuracy, sensitivity, and specificity values of 84.43%, 91.15%, and 79.16%, respectively.

Previous studies have shown that hybrid machine learning models improve predictive accuracy, robustness, and healthcare decision support by combining complementary techniques. However, challenges related to interpretability, feature weighting, uncertainty management, and effective decision fusion remain. To address these limitations, this study proposes a hybrid WCBA–BPNN–Fuzzy framework that integrates feature importance analysis, non-linear learning, and explainable reasoning to enhance predictive performance, interpretability, and clinical reliability in heart disease prediction.

3. DATASET SOURCE AND CHARACTERISTICS

The study utilised the Cleveland Heart Disease Dataset from the UCI Machine Learning Repository, a widely used benchmark dataset for cardiovascular disease prediction. The dataset contains 303 patient records with 13 predictive attributes and one target variable representing the presence or absence of heart disease. Its combination of demographic, physiological, and diagnostic features makes it suitable for developing and evaluating intelligent heart disease prediction models.

4. FEATURE DESCRIPTION AND CLINICAL RELEVANCE

The dataset incorporates thirteen clinically significant predictors that have been widely acknowledged as important risk indicators for cardiovascular disease. These variables represent a range of demographic, physiological, and diagnostic factors associated with cardiac health.

Table 1. Clinical Attributes Used in the Study

Feature	Description	Clinical Relevance
Age	Patient age (years)	Cardiovascular risk generally increases with advancing age
Sex	Male = 1, Female = 0	Males tend to exhibit a higher risk at earlier ages
Chest Pain Type (cp)	Four categories	Strongly associated with coronary artery abnormalities
Resting Blood Pressure (trestbps)	mmHg	Elevated blood pressure increases cardiac workload
Serum Cholesterol (chol)	mg/dL	High cholesterol contributes to atherosclerotic plaque formation
Fasting Blood Sugar (fbs)	>120 mg/dL (1/0)	Indicator of diabetic risk factors
Resting ECG (restecg)	Electrocardiographic results	Assists in identifying cardiac abnormalities
Maximum Heart Rate Achieved (thalach)	Beats per minute	Reflects cardiovascular fitness and function
Exercise-Induced Angina (exang)	Yes = 1, No = 0	Suggests restricted blood flow to the heart
ST Depression (oldpeak)	Continuous value	Indicates the severity of myocardial ischaemia
Slope of ST Segment (slope)	Three categories	Represents exercise-induced ECG characteristics
Number of Major Vessels (ca)	0–3	Higher values indicate increased likelihood of coronary disease
Thalassemia (thal)	Three categories	Associated with blood oxygen transport abnormalities

Among these variables, chest pain type, exercise-induced angina, ST depression, and the number of major vessels have consistently demonstrated strong predictive capability in previous cardiovascular studies. Consequently, these features were assigned greater importance during the WCBA-based feature weighting process incorporated within the proposed hybrid framework.

5. DATA CLEANING AND TREATMENT OF MISSING VALUES

To ensure data quality, missing values in the Cleveland dataset were identified and imputed using mean/median methods for numerical attributes and mode imputation for categorical attributes. Outlier analysis revealed no extreme observations requiring removal. The target variable was converted into a binary classification outcome, and duplicate records were removed where necessary. These pre-processing steps produced a clean and reliable dataset for model development.

6. FEATURE NORMALISATION AND ENCODING

As the dataset contains both numerical and categorical variables, additional pre-processing was required to ensure compatibility with machine learning algorithms, particularly the neural network component of the proposed framework.

Normalisation of Numerical Variables

All continuous attributes were normalised using Min–Max scaling according to Equation (1):

$$
X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}
$$

This transformation maps feature values to the interval [0,1], thereby reducing the influence of differing measurement scales. Normalisation contributes to faster convergence during neural network training, prevents variables with larger magnitudes from dominating the learning process, and improves the stability of gradient-based optimisation.

Encoding of Categorical Variables

Categorical features were converted into numerical representations suitable for computational processing. Ordinal variables were encoded using label encoding techniques, whilst nominal variables were transformed through one-hot encoding. Binary attributes, including sex, fasting blood sugar, and exercise-induced angina, were retained in their original binary format.

The resulting encoded dataset ensured appropriate representation of all features for both neural network training and association rule generation.

Training and Validation Strategy

To evaluate the generalisation capability of the proposed model, the dataset was divided into separate training and testing subsets. A conventional hold-out strategy was adopted, allocating 80% of the observations for training and the remaining 20% for independent testing.

In addition to the hold-out approach, stratified 10-fold cross-validation was employed to obtain a more robust assessment of predictive performance and minimise evaluation bias. Under this procedure, the dataset was randomly shuffled and partitioned into ten approximately equal subsets. During each iteration, nine folds were utilised for training and the remaining fold was reserved for testing. The process was repeated until every fold had served as the validation set once, after which the average performance metrics were computed.

The use of stratified sampling ensured that the proportion of positive and negative heart disease cases remained consistent across all folds and dataset partitions. This validation strategy enhances statistical reliability, reduces the likelihood of overfitting, and facilitates a fair comparison between standalone and hybrid predictive models.

7. PROPOSED HYBRID FRAMEWORK

The overall architecture of the proposed hybrid framework is illustrated in Figure 1.

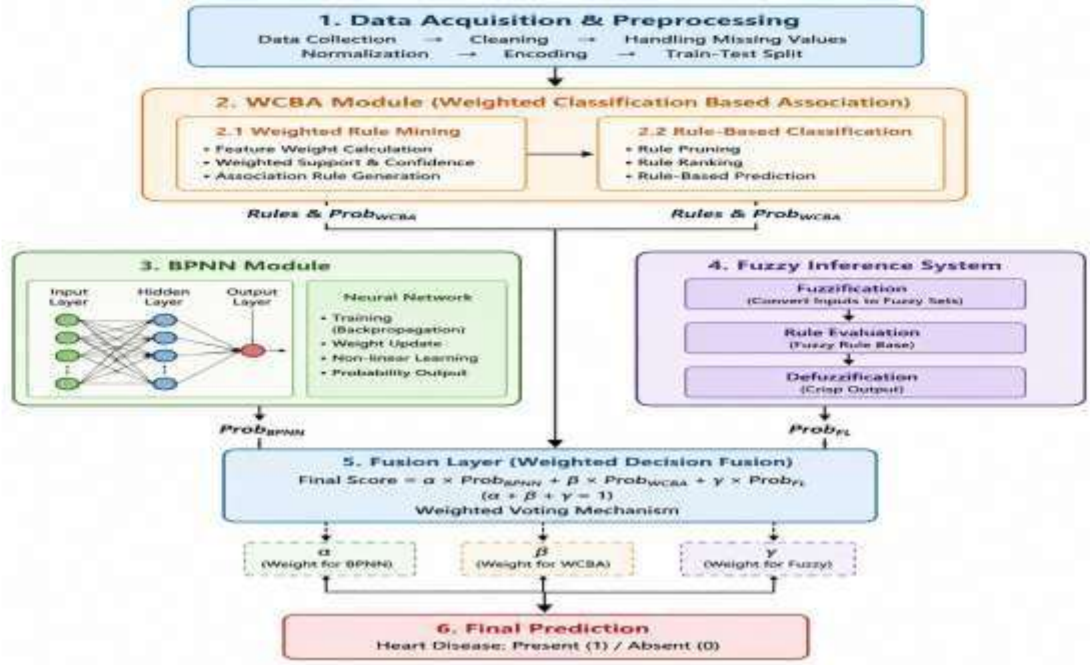

Fig. 1 Architecture of the Proposed Hybrid Heart Disease Prediction Model

The proposed hybrid framework integrates WCBA, BPNN, and Fuzzy Logic to address key challenges in heart disease prediction, including feature redundancy, non-linear relationships, and uncertainty in clinical data. WCBA performs feature prioritisation and rule extraction, BPNN learns complex patterns, and Fuzzy Logic provides interpretable reasoning. The combination of these complementary techniques enhances predictive accuracy, robustness, and clinical relevance.

Algorithm 1: Proposed Hybrid WCBA–BPNN–Fuzzy Framework for Heart Disease Prediction
Input: Pre-processed heart disease dataset (D)
Output: Final prediction outcome (P_{final})
1. Acquire and pre-process the dataset by performing data cleaning, missing value treatment, feature normalisation, and categorical feature encoding.
2. Apply the Weighted Classification Based Association (WCBA) module to determine feature importance and generate weighted class association rules.
3. Select the most informative weighted features and compute the WCBA prediction probability ((P_{WCBA})).
4. Train the Back-Propagation Neural Network (BPNN) using the selected weighted features and obtain the neural network prediction probability ((P_{BPNN})).
5. Perform fuzzy reasoning by applying the Fuzzy Inference System (FIS) to the selected clinical attributes together with the neural network output, producing a fuzzy prediction score ((P_{Fuzzy})).
6. Integrate the outputs of the three modules through a weighted decision fusion mechanism according to:
[
$$P_{final}=\alpha P_{BPNN}+\beta P_{WCBA}+\gamma P_{Fuzzy}$$
]
subject to the constraint:
[
$$\alpha+\beta+\gamma=1$$
]
where (α), (β), and (γ) represent the contribution weights of the BPNN, WCBA, and Fuzzy modules, respectively.
7. Compare the fused prediction score ((P_{final})) with a predefined decision threshold ((τ)).
8. Assign the corresponding class label:
o Heart Disease Present, if ($P_{final} \geq \tau$)
o Heart Disease Absent, if ($P_{final} < \tau$)
9. Return the final diagnostic prediction.

Weighted Classification Based Association (WCBA) Layer
The WCBA layer performs feature prioritisation and rule generation by assigning weights according to the clinical significance of individual attributes. Clinically important predictors, such as chest pain type, ST-segment depression, maximum heart rate, and exercise-induced angina, are given higher importance during learning. The approach generates high-confidence class association rules, removes less informative features, and provides interpretable IF–THEN rules, thereby enhancing both predictive performance and transparency in heart disease prediction.

Back-Propagation Neural Network (BPNN) Layer
The BPNN layer models complex non-linear relationships among clinical variables involved in heart disease prediction. By learning adaptive weights through gradient-based optimisation, the network identifies hidden patterns and interactions that may not be captured by rule-based methods. This enhances the predictive capability, classification accuracy, and generalisation performance of the proposed hybrid framework.

Fuzzy Logic Inference Layer
The Fuzzy Logic layer manages uncertainty in clinical decision-making by converting numerical inputs into linguistic categories using membership functions and expert-defined fuzzy rules. Through Mamdani inference, it generates interpretable risk assessments, enhancing the transparency, reliability, and clinical relevance of the final prediction.

Decision Fusion Mechanism

The final prediction is generated through a weighted fusion of the WCBA, BPNN, and Fuzzy Logic outputs. By combining rule-based reasoning, non-linear learning, and uncertainty management, the fusion mechanism improves robustness, sensitivity, specificity, and generalisation performance. This integrated approach provides a more reliable and effective diagnostic solution than any individual technique alone.

7. RESULTS AND PERFORMANCE EVALUATION

The effectiveness of the proposed hybrid heart disease prediction framework was assessed using widely accepted classification performance measures, namely Accuracy, Precision, Recall (Sensitivity), F1-Score, Specificity, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). Experimental evaluation was conducted using the Cleveland Heart Disease Dataset obtained from the UCI Machine Learning Repository. To ensure reliable performance estimation and minimise bias arising from data partitioning, stratified 10-fold cross-validation was employed throughout the experimentation process.
This section presents a comparative evaluation of the Hybrid BPNN–CBA–Fuzzy model and the proposed BPNN–WCBA–Fuzzy framework to assess the effectiveness of incorporating weighted association rule mining within the hybrid architecture. The performance of both models is evaluated using standard classification metrics, including accuracy, precision, recall, F1-score, and AUC. Furthermore, confusion matrix analysis and other performance indicators are utilised to examine the classification behaviour, diagnostic reliability, and predictive effectiveness of the proposed framework.

Comparative Analysis of Hybrid Models

To investigate the effectiveness of incorporating weighted association rule mining within the hybrid architecture, a comparative analysis was performed between the Hybrid BPNN–CBA–Fuzzy model and the proposed BPNN–WCBA–Fuzzy framework. Both models combine neural learning and fuzzy reasoning; however, the proposed framework replaces conventional Classification Based on Association (CBA) with Weighted Classification Based on Association (WCBA), enabling feature importance to be incorporated during the classification process.

Table 2 presents the comparative performance of the two hybrid models.

Model	Accuracy (%)	Precision	Recall	F1-Score	AUC
Hybrid BPNN–CBA–Fuzzy	89.40	0.82	0.90	0.88	0.91
Proposed BPNN–WCBA–Fuzzy	92.08	0.91	0.93	0.92	0.95

The proposed BPNN–WCBA–Fuzzy framework outperformed the conventional BPNN–CBA–Fuzzy model across all evaluation metrics. It achieved an accuracy of 92.08% and an AUC of 0.95, demonstrating improved predictive performance and diagnostic reliability. The incorporation of WCBA enhanced feature relevance assessment, reduced classification errors, and improved the identification of heart disease cases. These findings confirm that integrating weighted association rule mining strengthens the hybrid architecture, making it a more effective and clinically applicable solution for heart disease prediction.

8. LIMITATIONS OF THE PROPOSED FRAMEWORK

Despite its promising performance, the proposed WCBA–BPNN–Fuzzy framework has limitations, including evaluation on a single dataset, increased computational complexity, and dependence on optimal hyperparameter selection. Additionally, the neural network component remains less interpretable, and the model has not yet been validated in real-world clinical settings. Therefore, it should be considered a decision-support tool rather than a substitute for clinical expertise.

9. FUTURE RESEARCH DIRECTIONS

Future work will focus on validating the proposed framework using larger and more diverse heart disease datasets to improve its generalisability and robustness. The integration of advanced deep learning techniques, such as CNNs and LSTMs, may further enhance predictive performance.

Additionally, Explainable AI methods, including SHAP and LIME, can be incorporated to improve transparency and clinician trust. The framework may also be extended to real-time clinical decision-support systems and adapted for the prediction of other chronic diseases, such as diabetes and kidney disease.

10. CONCLUSION

This study proposed a novel hybrid WCBA–BPNN–Fuzzy framework for heart disease prediction, integrating weighted association rule mining, neural network learning, and fuzzy reasoning. Experimental results demonstrated superior performance over conventional approaches, achieving an accuracy of 92.08% and an AUC of 0.95. Comparative analysis confirmed that the proposed framework outperformed the conventional BPNN–CBA–Fuzzy model, highlighting the effectiveness of weighted feature selection and decision fusion. Furthermore, the framework enhances clinical interpretability and uncertainty handling, making it suitable for intelligent decision-support applications. Future research will focus on validation using larger datasets and the integration of Explainable AI techniques to further improve transparency and clinical applicability.

REFERENCES

[1] World Health Organization, *Cardiovascular Diseases (CVDs)*, Geneva, Switzerland: WHO, 2021.

[2] D. Dua and C. Graff, "UCI Machine Learning Repository," University of California, Irvine, CA, USA, 2019.

[3] K. Polat and S. Güneş, "A hybrid approach to heart disease classification," *Expert Systems with Applications*, vol. 38, no. 7, pp. 7904–7910, 2011.

[4] V. Vapnik, *Statistical Learning Theory*. New York, NY, USA: Wiley, 1998.

[5] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[6] D. Saha et al., "Heart disease prediction using logistic regression," in *Proceedings of the 2025 International Conference on Computer, Electrical and Communication Engineering (ICCECE)*, Kolkata, India, 2025, pp. 1–6, doi: 10.1109/ICCECE61355.2025.10940073.

[7] B. Duraisamy, R. Sunku, K. Selvaraj, V. V. R. Pilla, and M. Sanikala, "Heart disease prediction using support vector machine," *Multidisciplinary Science Journal*, vol. 6, p. 2024ss0104, 2023, doi: 10.31893/multiscience.2024ss0104.

[8] U. M. S. and L. Joshitha, "Heart disease prediction using decision tree variants: A comparative analysis," in *Proceedings of the 2025 International Conference on Computing and Communication Technologies (ICCCT)*, Chennai, India, 2025, pp. 1–6, doi: 10.1109/ICCCT63501.2025.11019035.

[9] R. R. Devi, P. Dharshini, R. Hemala, and D. Swetha, "Heart disease prediction using random forest classifier," in *Proceedings of the 2023 International Conference on Data Science, Agents and Artificial Intelligence (ICDSAAI)*, Chennai, India, 2023, pp. 1–8, doi: 10.1109/ICDSAAI59313.2023.10452459.

[10] S. Ghouhar Taj and K. Kalaivani, "Heart disease prediction using KNN algorithm approach," *ARPN Journal of Engineering and Applied Sciences*, vol. 19, no. 12, pp. 754–764, Jun. 2024.

[11] A. A. Shinde, S. N. Kale, R. M. Samant, A. S. Naik, and S. A. Ghorpade, "Heart disease prediction system using multilayered feed-forward neural network and back-propagation neural network," *International Journal of Computer Applications*, vol. 166, no. 7, pp. 32–36, May 2017, doi: 10.5120/ijca2017914080.

[12] N. Sabri et al., "HeartInspect: Heart disease prediction of an individual using naïve Bayes algorithm," in *Proceedings of the IEEE 11th Conference on Systems, Process and Control (ICSPC)*, Malacca, Malaysia, 2023, pp. 350–354, doi: 10.1109/ICSPC59664.2023.10420149.

[13] F. D. M. Inamdar, N. JayaKumar, and B. Bari, "Heart disease predictive analysis using association rule mining," in *Intelligent Communication Technologies and Virtual Mobile Networks*, Singapore: Springer, 2023, pp. 1–10, doi: 10.1007/978-981-19-1844-5_10.

[14] A. El-Ibrahimi, O. Daanouni, Z. Alouani, O. El Gannour, S. Saleh, B. Cherradi, and O. Bouattane, "Fuzzy-based system for coronary artery disease prediction using subtractive clustering and risk factor data," *Intelligence-Based Medicine*, vol. 11, 2025, doi: 10.1016/j.ibmed.2025.100208.

[15] M. Bhuiyan, C. Wah, N. Kamarudin, N. Ismail, and A. F. Ab. Nasir, "Heart disease prediction using case-based reasoning (CBR)," *Journal of Theoretical and Applied Information Technology*, vol. 102, no. 20, pp. 7275–7286, 2025.

[16] A. Diwan, S. Karmakar, and S. Soni, "Design and development of a machine learning-based hybrid model using BPNN and associative classifier for heart disease prediction," *MSW Management Journal*, vol. 35, no. 2, 2025, doi: 10.7492/vkf74q14.

[17] J. Kaur and B. Khehra, "Fuzzy logic and hybrid-based approaches for the risk of heart disease detection: State-of-the-art review," *Journal of the Institution of Engineers (India): Series B*, vol. 103, no. 2, pp. 681–697, 2021.

[18] M. Aghamohammadi, M. Madan, J. K. Hong, and I. Watson, "Predicting heart attack through explainable artificial intelligence," in *Computational Science – ICCS*, vol. 11537, pp. 633–645, 2019.