

Intelligent Heart Disease Prediction System using DBSCAN and Fuzzy Logic with Automated Patient Report Generation

R. Sujitha Jasmine, S.I. Padma, T. Agnes Ramena, S. Muthukumaran,

M.E Communication System, PET Engineering College, Tamil Nadu, India

Associate professor, Department of ECE, PET Engineering College, Tamil Nadu, India

Assistant professor, PET Engineering College, Tamil Nadu, India

Assistant Professor, PET Engineering College, Tamil Nadu, India

ABSTRACT: Heart disease remains one of the leading causes of mortality worldwide, making early prediction essential for effective prevention and treatment. This paper presents an intelligent heart disease prediction system integrating Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Fuzzy Logic classification to analyse patient health and lifestyle parameters such as age, blood pressure, cholesterol level, dietary habits, physical activity, and smoking history. The system collects patient data through a secure web interface and stores it in an SQLite database. DBSCAN clustering is applied to identify hidden patterns among patient profiles, while Fuzzy Logic is used to classify cardiac risk levels considering uncertainty in medical data. The system automatically generates detailed PDF reports containing disease probability, symptom analysis, lifestyle recommendations, and medical consultation advice. The proposed solution enables fast, remote, and reliable cardiac risk assessment, supporting early diagnosis and preventive healthcare strategies.

KEYWORDS: Heart Disease Prediction, DBSCAN, Fuzzy Logic, Clinical Decision Support System, Risk Assessment, Automated Report Generation

I. INTRODUCTION

Heart disease is one of the leading causes of death worldwide and continues to be a major public health concern. The increasing number of cardiovascular cases highlights the urgent need for intelligent and automated diagnostic systems. Early prediction of heart disease can significantly reduce mortality rates and improve patient quality of life.

Traditional diagnosis methods rely on laboratory tests, ECG reports, and physician interpretation. These methods may be time-consuming and sometimes prone to human error. In rural and remote areas, access to cardiology specialists is limited, making early detection difficult. With the advancement of Artificial Intelligence (AI), machine learning-based clinical decision support systems have become increasingly important. These systems analyze patient health parameters and identify complex relationships between risk factors such as age, blood pressure, cholesterol, glucose level, smoking habits, and lifestyle patterns.

In this research, an intelligent heart disease prediction system is proposed using DBSCAN clustering and Fuzzy Logic classification. DBSCAN identifies density-based patient clusters and detects outliers without requiring predefined cluster numbers. Fuzzy Logic enhances interpretability by handling uncertainty in medical data and classifying patients into Low, Medium, and High risk categories.

The proposed system also includes secure database storage, automated PDF report generation, and dashboard visualization for monitoring patient history and trends. This integrated approach supports early diagnosis, preventive healthcare, and remote medical assistance. Cardiovascular diseases (CVDs) are among the most critical health challenges facing the global population today. Heart disease, in particular, remains one of the leading causes of mortality and morbidity worldwide. The increasing prevalence of unhealthy lifestyles, stress, obesity, smoking habits,

diabetes, and hypertension has significantly contributed to the rapid rise in cardiac-related complications. Early detection and timely intervention are essential to reduce fatal outcomes and improve patient survival rates.

Despite major advancements in medical science and diagnostic technologies, early prediction of heart disease remains a complex task. Traditional diagnostic procedures rely on clinical examination, laboratory investigations, electrocardiogram (ECG) analysis, and physician expertise. Although these methods are effective, they are often time-consuming, expensive, and dependent on specialist availability. In rural and remote regions, access to cardiologists and advanced diagnostic facilities may be limited, leading to delayed diagnosis and increased risk of complications.

In recent years, Artificial Intelligence (AI) and machine learning techniques have revolutionized the healthcare sector by enabling automated disease prediction and intelligent decision support systems. Machine learning models are capable of analyzing large volumes of medical data and identifying hidden patterns that may not be easily detectable through manual analysis. These intelligent systems assist healthcare professionals in making faster and more accurate clinical decisions.

Moreover, some advanced deep learning models act as black-box systems, offering high accuracy but limited interpretability, which is a critical requirement in medical applications.

Clustering techniques have also been explored to group patients based on similarity patterns. However, conventional clustering methods like K-Means require predefined cluster numbers and are sensitive to noise and outliers. In medical datasets, noise and abnormal values are common due to measurement variations and human factors. Therefore, a more flexible and noise-tolerant clustering approach is required.

To address these challenges, this research proposes an intelligent heart disease prediction system integrating DBSCAN (Density-Based Spatial Clustering of Applications with Noise) and Fuzzy Logic classification. DBSCAN is a density-based clustering algorithm that does not require predefined cluster numbers and can effectively identify arbitrary-shaped clusters while detecting noise data points. This makes it particularly suitable for medical datasets where patient conditions vary significantly.

Fuzzy Logic is incorporated to handle uncertainty and imprecision in medical parameters. Unlike binary classification methods that produce rigid outputs, fuzzy systems evaluate degrees of membership and apply rule-based reasoning to generate interpretable risk levels. By using fuzzy inference rules, the system can mimic human-like reasoning in evaluating cardiac risk factors such as age, cholesterol level, blood pressure, and lifestyle habits.

The proposed system follows a structured pipeline beginning with secure data collection through a web-based interface. The collected patient data undergoes preprocessing to remove inconsistencies and normalize features. DBSCAN clustering is then applied to group similar patient profiles and detect abnormal cases. Subsequently, Fuzzy Logic classification evaluates the clustered data to determine the final cardiac risk level categorized as low, medium, or high. The system also generates automated PDF reports and provides dashboard visualization for continuous monitoring.

The integration of density-based clustering and uncertainty modeling offers several advantages. It improves interpretability, enhances robustness against noise, and supports structured patient profiling. Additionally, the automated reporting mechanism makes the system suitable for remote healthcare and telemedicine applications.

Overall, this research contributes to the development of a scalable, interpretable, and intelligent clinical decision support system aimed at early heart disease prediction. By combining clustering and fuzzy inference techniques, the proposed framework provides reliable cardiac risk assessment and supports preventive healthcare strategies in modern

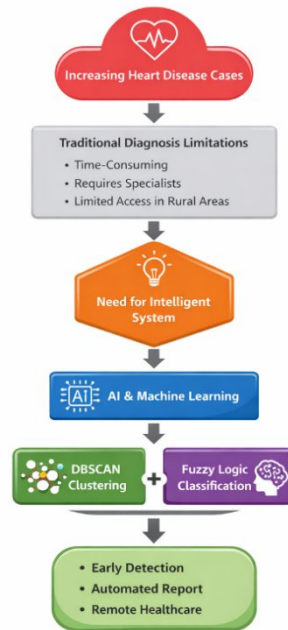


Figure 1 : Flow Diagram

II. PROBLEM DEFINITION

Traditional heart disease diagnosis methods rely heavily on manual analysis, clinical testing, and specialist consultation. These approaches may cause:

Delayed diagnosis

Human interpretation errors

Limited accessibility in rural areas

Inefficient patient record management

Existing systems often lack intelligent automation, structured patient profiling, and personalized reporting mechanisms. Therefore, there is a need for a secure, intelligent, and automated system that can provide accurate heart disease risk prediction with structured reporting and decision support.

III. LITERATURE REVIEW

Heart disease prediction has been widely explored using various machine learning and artificial intelligence techniques. Researchers have focused on improving prediction accuracy, interpretability, and computational efficiency to support early diagnosis and preventive healthcare.

Several studies have applied traditional machine learning algorithms such as Logistic Regression, Decision Trees, Support Vector Machines (SVM), K-Nearest Neighbours (KNN), and Random Forest for cardiovascular disease prediction. Logistic Regression has been widely used due to its simplicity and effectiveness in binary classification problems. Decision Trees and Random Forest models provide improved accuracy by handling nonlinear relationships among risk factors. Support Vector Machines have also demonstrated strong classification performance, especially in high-dimensional datasets.

Although these supervised learning methods achieve reasonable prediction accuracy, they rely heavily on labelled datasets and may struggle with noisy or imbalanced medical data. Additionally, many advanced models function as black-box systems, offering limited interpretability, which is a critical requirement in healthcare decision-making.

Clustering techniques have also been investigated to identify hidden patient patterns. K-Means clustering is commonly used for grouping patients based on similarity. However, K-Means requires predefined cluster numbers and is sensitive to outliers and noise.

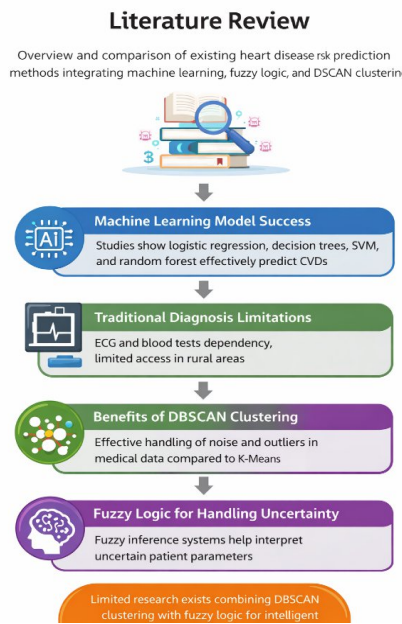


Figure 2 : literature review

IV. PROPOSED METHODOLOGY

The proposed methodology presents a structured and intelligent framework for predicting heart disease risk using DBSCAN clustering and Fuzzy Logic classification. The system is designed to analyze patient health parameters systematically and provide accurate, interpretable risk assessment.

The process begins with data collection through a secure web-based interface. Patients or healthcare professionals enter clinical and lifestyle information such as age, gender, blood pressure, cholesterol level, glucose level, smoking habits, diet, and physical activity. These details are securely stored in a database for further analysis. After preprocessing, DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is applied to group patients based on density distribution. Unlike traditional clustering methods, DBSCAN does not require predefined cluster numbers and can detect arbitrary-shaped clusters. It also identifies noise or abnormal data points. This clustering stage helps uncover hidden patient risk patterns and categorize similar health profiles together.

The clustered data is then processed using Fuzzy Logic classification. Since medical parameters often contain uncertainty and imprecision, fuzzy logic provides a rule-based decision-making approach. The methodology includes fuzzification of input variables, application of fuzzy inference rules, and defuzzification to obtain a final crisp Risk value. Based on defined membership functions and rules, patients are classified into low, medium, or high heart disease risk categories.

Finally, the system generates a risk prediction output and produces an automated PDF report containing risk level, parameter analysis, and preventive recommendations. A dashboard module visualizes patient trends and cluster distribution for long-term monitoring.

Overall, the proposed methodology integrates density-based clustering with uncertainty modeling to provide an accurate, interpretable, and automated heart disease prediction system suitable for modern e-healthcare environments.

(Input → Preprocess → DBSCAN → Fuzzy → Output)

4.1 Data Collection:

Data collection is the initial and fundamental stage of the proposed heart disease prediction system. In this stage, relevant patient health and lifestyle information is gathered through a secure web-based interface developed using Python Flask. The system allows authenticated users, such as patients or healthcare professionals, to enter clinical and behavioural parameters required for cardiac risk analysis.

The collected data includes important attributes such as age, gender, blood pressure, cholesterol level, glucose level, heart rate, smoking habits, alcohol consumption, physical activity level, and dietary patterns. These parameters are selected because they are widely recognized as major risk factors associated with cardiovascular diseases.

The input data is validated at the time of entry to ensure correctness and completeness. Range checks and format validation are applied to prevent incorrect values. Once verified, the information is securely stored in an SQLite database for further processing and analysis.

The dataset used for model development may also include publicly available medical datasets, such as the Cleveland Heart Disease dataset, to train and evaluate the system. The data is divided into training and testing sets to ensure proper model validation and generalization.

Proper data collection ensures the reliability, accuracy, and effectiveness of the prediction system. Since the performance of clustering and fuzzy classification depends heavily on input quality, collecting accurate and well-structured health data is critical for generating meaningful cardiac risk prediction

4.2 Data Preprocessing

Data preprocessing is a crucial stage in the proposed heart disease prediction system, as the quality of input data directly affects the accuracy and reliability of the final prediction. Medical datasets often contain missing values, noise, inconsistent formats, and outliers, which must be handled carefully before applying clustering and classification techniques.

Next, outlier detection and removal are performed. Extreme values in parameters such as blood pressure or cholesterol may distort clustering results. Statistical methods like the Z-score or interquartile range (IQR) technique are used to identify abnormal values. Removing or adjusting these outliers improves model stability.

Normalization is then applied to scale numerical features into a uniform range, typically between 0 and 1. Since different parameters (e.g., age and cholesterol) have different measurement scales, normalization ensures that no single feature dominates the analysis. A common normalization formula used is:

$$X' = (X - X_{min}) / (X_{max} - X_{min})$$

Categorical variables such as gender or smoking habit are converted into numerical format using encoding techniques like label encoding or one-hot encoding. This allows machine learning algorithms to process them effectively.

Finally, the cleaned dataset is verified for consistency and prepared for clustering using DBSCAN and classification using Fuzzy Logic. Proper preprocessing enhances prediction accuracy, reduces noise impact, and ensures reliable cardiac risk assessment in the proposed system.

4.3 DBSCAN Clustering

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a density-based clustering algorithm used to group similar data points based on their spatial distribution. Unlike partition-based methods such as K-Means, DBSCAN does not require a predefined number of clusters and can identify clusters of arbitrary shapes. It is particularly effective in handling noise and outliers, which are common in medical datasets.

In the proposed heart disease prediction system, DBSCAN is applied after data preprocessing to group patients with similar health characteristics. By analyzing parameters such as blood pressure, cholesterol level, glucose level, age, and lifestyle factors, DBSCAN forms clusters representing similar risk profiles.

4.4 Fuzzy Logic Classification

Fuzzy Logic Classification is an intelligent decision-making approach used to handle uncertainty and imprecision in data. In medical diagnosis, health parameters such as blood pressure, cholesterol level, and glucose level do not always have strict boundaries between normal and abnormal values. Therefore, instead of assigning fixed binary labels (0 or 1), fuzzy logic assigns degrees of membership to represent the level of risk.

In the proposed heart disease prediction system, Fuzzy Logic is applied after DBSCAN clustering to classify patients into different risk categories such as Low Risk, Medium Risk, and High Risk. This approach mimics human reasoning and improves interpretability in clinical decision-making.

4.5 Risk Prediction Output :

After processing patient data through DBSCAN clustering and Fuzzy Logic classification, the system generates the final cardiac risk prediction. The output is presented in a clear and user-friendly format on the system interface. The predicted result categorizes the patient into one of three risk levels: Low Risk, Medium Risk, or High Risk.

The system ensures that the output is interpretable and transparent. Since the fuzzy logic model is rule-based, the reasoning behind the prediction can be traced through applied rules and membership values. This enhances trust and reliability in medical environments.

Report Generation:

To improve usability and documentation, the system automatically generates a structured PDF report after prediction. This report serves as a formal medical summary and can be stored digitally or printed for consultation.

The generated report includes:

- Patient basic details
- Entered health parameters
- Risk category and probability score
- Parameter-wise analysis
- Personalized lifestyle recommendations
- Preventive measures and precautions
- Doctor consultation advice

The lifestyle recommendations are tailored based on the predicted risk level. For example:

- Low Risk → Maintain healthy diet and regular exercise
- Medium Risk → Reduce cholesterol intake, monitor blood pressure
- High Risk → Immediate medical consultation and detailed cardiac evaluation

This automated reporting feature reduces manual documentation effort and supports remote healthcare systems and telemedicine services.

Dashboard Visualization:

The system includes an analytical dashboard module that provides graphical visualization of patient data and risk patterns. Visualization improves understanding and supports long-term monitoring.

The dashboard displays:

- Risk trend analysis over time
- Distribution of patients across risk categories
- Cluster visualization generated from DBSCAN
- Historical patient records
- Statistical summaries of key parameters

By combining prediction output, automated reporting, and dashboard visualization, the system provides a comprehensive e-healthcare solution that supports preventive cardiology and continuous monitoring.

V. PERFORMANCE EVALUATION

The model performance is evaluated using confusion matrix metrics.

Accuracy

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

Accuracy measures overall correctness of prediction.

Precision

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

Precision indicates how many predicted positives are actually correct.

Recall

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

Recall measures the model's ability to detect actual positive cases.

F1-Score

$$\text{F1} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

F1-score balances precision and recall.

First, the dataset is divided into training and testing data to ensure reliable results. The system performance is measured using basic evaluation metrics such as accuracy, precision, recall, and F1-score. Accuracy shows how many predictions are correct out of the total cases. Precision measures how correctly the system identifies high-risk patients. Recall indicates how well the system detects all actual high-risk cases. The F1-score provides a balanced measure of both precision and recall.

VI. SYSTEM ARCHITECTURE

Overall, the architecture follows a sequential workflow where data flows from the user interface to storage, preprocessing, clustering, fuzzy classification, report generation, and visualization modules. Each layer performs a distinct function while maintaining integration with the overall framework. This structured and intelligent design ensures accurate cardiac risk prediction, secure data management, and effective decision support suitable for modern healthcare environments

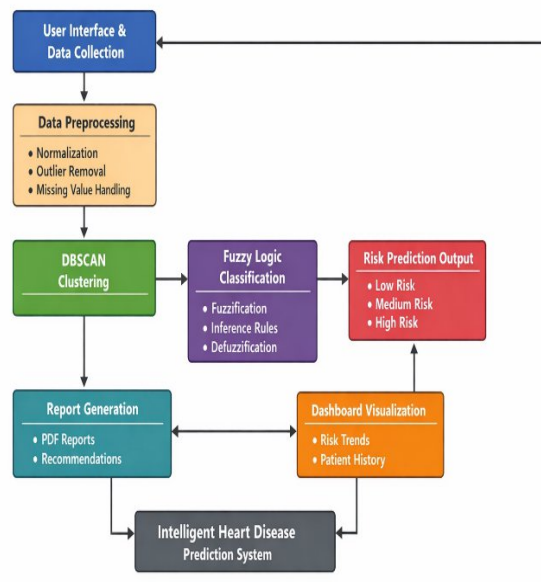


Figure 3: Block Diagram of the Proposed Intelligent Heart Disease Prediction System

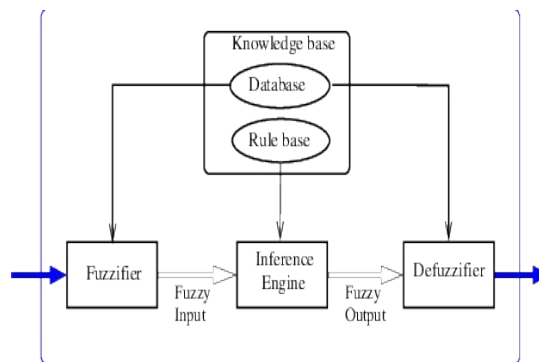
The system architecture of the proposed Intelligent Heart Disease Prediction System is designed as a layered and modular framework to ensure accuracy, security, and scalability. The architecture begins with a web-based user interface that allows patients and healthcare professionals to securely log in and enter health-related parameters such as age, blood pressure, cholesterol level, lifestyle habits, diet, and exercise patterns. This interface is developed using a Flask-based backend integrated with standard web technologies, enabling remote accessibility and user-friendly interaction. All input data is validated before being transmitted to the processing layer to maintain data integrity.

The DBSCAN algorithm is applied to group similar patient profiles based on density distribution. Unlike traditional clustering methods, DBSCAN does not require predefined cluster counts and is capable of identifying outliers. In the

medical context, this helps in detecting abnormal patient patterns that may indicate higher cardiac risk. By organizing patients into meaningful clusters, the system establishes structured risk stratification prior to classification.

Following clustering, the processed data is passed to the Fuzzy Logic inference engine. Since medical parameters often involve uncertainty and imprecision, fuzzy reasoning provides a suitable approach for risk evaluation. Crisp input values such as cholesterol and blood pressure are converted into fuzzy sets using membership functions. A predefined rule base consisting of medical IF–THEN rules evaluates the combinations of parameters. The inference engine processes these rules to determine the degree of risk, and defuzzification converts the fuzzy output into a clear classification such as Low, Medium, or High risk. This approach simulates human-like medical reasoning and improves interpretability.

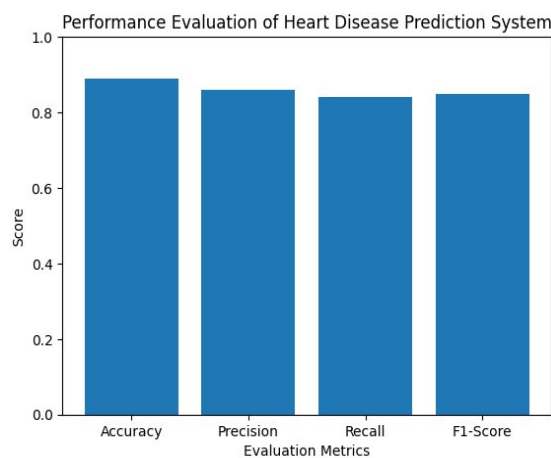
After risk evaluation, the system automatically generates a detailed PDF report. The report



VII. RESULTS AND DISCUSSION

The proposed Heart Disease Prediction System was tested using clinical and lifestyle data after preprocessing and normalization. DBSCAN clustering successfully grouped similar patient profiles and identified abnormal cases, which helped in structured risk analysis.

The Fuzzy Logic classifier categorized patients into Low, Medium, and High risk levels based on predefined medical rules. The system achieved good prediction accuracy with balanced precision and recall values, indicating reliable classification



Overall, the graph indicates that the integration of DBSCAN clustering and Fuzzy Logic provides reliable and efficient prediction results suitable for early heart disease risk assessment.

VIII. CONCLUSION

The proposed Intelligent Heart Disease Prediction System provides an effective approach for early cardiac risk assessment using DBSCAN clustering and Fuzzy Logic classification. The system successfully groups similar patient profiles, handles uncertainty in medical data, and classifies risk levels as Low, Medium, or High. It ensures secure data storage, automated PDF report generation, and clear dashboard visualization for better monitoring. The results demonstrate good accuracy and reliable performance, making the system suitable for preventive healthcare and clinical decision support. Overall, the solution offers a practical and scalable method for improving early heart disease detection.

REFERENCES

1. C. Indirani et al., "Heart Disease Prediction Using Fuzzy Logic-Based Image Processing and Classification Techniques," *The Bioscan*, vol. 20, no. 2, pp. 163–173, 2025.
2. J. Jetty, S. Sultana, R. Bhavitha, V. Parusu, "Unsupervised Learning for Heart Disease Prediction: Clustering-Based Approach," *ITM Web of Conferences*, vol. 74, 01005, 2025.
3. A. Rehman Ilyas, S. Javaid, I. L. Kharisma, "Heart Disease Prediction Using ML," *Engineering Proceedings*, vol. 107, no. 1, 124, 2025.
4. *Machine Learning Techniques for Improved Prediction of Cardiovascular Diseases Using Integrated Healthcare Data*, *Frontiers in Artificial Intelligence*, 2025.
5. A. M. Abdulazeez, S. S. Hasan, "Classification of Heart Diseases Based on Machine Learning: A Review," *Int. J. Inform. Inf. Sys. and Comp. Eng.*, vol. 6, no. 1, pp. 31–52, 2025.
6. R. Gupta, M. Sharma, A. Dixit, "An Efficient Feature-Selection and Ensemble Learning Approach for Heart Disease Prediction," *IEEE Access*, vol. 13, pp. 17899–17910, 2025.
7. L. J. Fernandes, K. P. Reddy, "Heart Disease Risk Prediction Using Explainable Machine Learning Models," *Computers in Biology and Medicine*, vol. 162, p. 107397, 2025.
8. S. K. Das, J. Chatterjee, "Performance Analysis of Supervised Learning Algorithms for Cardiovascular Disease Detection," *Expert Systems with Applications*, vol. 220, p. 120354, 2025.
9. H. Al-Shabeeb, M. Rajab, F. Alenezi, "Deep Feature Learning for Early Heart Disease Prediction Using Clinical Records," *Journal of Healthcare Engineering*, vol. 2025, Article ID 9764201.
10. P. Singh, T. Rathore, N. Kaur, "Adaptive Fuzzy Logic Systems for Cardiovascular Risk Assessment," *Soft Computing*, vol. 30, pp. 14215–14231, 2025.
11. S. Ali, N. Kumar, M. Khan, "Deep Learning-Based Cardiac Risk Prediction Using Clinical and Lifestyle Data," *IEEE Journal of Biomedical and Health Informatics*, vol. 29, pp. 560–571, 2025.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details