Journal of Information Systems Engineering and Management
2025, 10(19s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

A Novel Deep Fuzzy Rule-Based System for Early Heart Disease Risk Prediction

G. Angayarkanni [1*], Dr. M. Rajasenathipathi[2]
[1]*Assistant Professor, Department of Computer Science, Nallamuthu Gounder Mahalingam College, Pollachi, Tamilnadu*
[2]*Associate Professor. Department of Computer Technology, Nallamuthu Gounder Mahalingam College, Pollachi, Tamilnadu*
** Corresponding Author: g.angayarkanni@gmail.com*

ARTICLE INFO

Received: 23 Dec 2024

Revised: 31 Jan 2025

Accepted: 15 Feb 2025

ABSTRACT

Heart disease continues to be a major factor in global deaths, highlighting the necessity for prediction models that maintain both accuracy and interpretability. This study introduces an innovative Deep Fuzzy Rule-Based Framework aimed at predicting heart disease, combining the analytical strengths of deep learning with the clarity provided by fuzzy logic. The framework utilizes deep learning for extracting features and acquiring representations, thus allowing for the detection of intricate patterns within extensive medical datasets. Fuzzy logic contributes to interpretability by producing rules that are understandable to humans, aiding in clinical decision-making. The methodology proposed encompasses preprocessing steps like handling missing values and normalization, fuzzy entropy-based feature selection for reducing dimensionality, and dynamic rule pruning to enhance computational efficiency. The framework performed quite well, achieving 92.1% accuracy, 93.3% sensitivity, 91.2% specificity, and an F1-score of 91.8% when tested on a massive dataset of 10 lakh cases. The findings support its scalability and stability, which makes it a useful clinical decision support tool for early heart disease detection.

Keywords: Heart disease prediction, mean imputation, fuzzy logic, deep learning, rule-based systems, big data analytics, fuzzy entropy.

1. INTRODUCTION

Heart disease ranks among the most prevalent causes of death globally, claiming millions of lives each year. Identifying the condition early and taking timely action is vital to reducing its toll. In recent years, advancements in artificial intelligence (AI) have paved the way for more sophisticated predictive models. However, challenges persist. Black-box AI models, like deep learning, often fall short in interpretability, a critical factor for clinical settings where understanding the reasoning behind predictions is paramount. On the other hand, rule-based approaches such as fuzzy logic provide transparency but struggle to handle the complexity and vastness of high-dimensional medical datasets.

This paper proposes a novel Deep Fuzzy Rule-Based Framework that bridges this gap, offering both interpretability and scalability. By merging the pattern-recognition strengths of deep learning with the human-readable reasoning of fuzzy logic, this framework aims to enhance heart disease prediction. Deep learning enables the automated extraction of features from extensive datasets, while fuzzy logic generates interpretable rules that align with medical expertise.

With the rise of big data, the healthcare industry now has access to enormous repositories of patient records, electronic health records (EHRs), and imaging data. Deep learning has demonstrated its effectiveness in analyzing these datasets, uncovering complex patterns that were previously undetectable [1]. However, clinical integration demands that these findings be interpretable, which is where fuzzy logic shines. Research by Ali et al. [2] and Bahani et al. [3] has shown the potential of hybrid models that combine fuzzy logic with machine learning for disease prediction. Similarly, Hameed et al. utilized fuzzy rules optimized via genetic algorithms to improve diagnostic accuracy, emphasizing the significance of hybrid approaches in managing medical data complexity. This study builds upon such advancements by evaluating the proposed framework on a large-scale dataset comprising one million

records. By incorporating big data analytics, it ensures scalability, while fuzzy logic addresses the need for interpretability. This framework strikes a balance between accuracy and decision-making transparency, making it a robust tool for real-world clinical applications.

2. OBJECTIVES

- Propose a novel Deep Fuzzy Rule-Based Framework for heart disease prediction that integrates deep learning for feature extraction and fuzzy logic for interpretability, addressing the dual need for accuracy and clinical decision support.
- Evaluate the framework on a large-scale dataset comprising one million instances to ensure scalability and robustness while achieving high performance metrics such as accuracy, sensitivity, and specificity for reliable heart disease diagnosis.
- Develop a practical clinical decision support system (CDSS) that balances predictive accuracy with human-understandable rules, improving interpretability and supporting healthcare professionals in early heart disease diagnosis and decision-making.

3. RELATED WORK

This literature review examines the use of fuzzy logic (FL) and hybrid intelligent systems in diagnosing and predicting heart disease, emphasizing recent developments. The analyzed studies reveal an increasing trend of combining FL with machine learning, neural networks, genetic algorithms, and other methodologies to enhance diagnostic precision, manage the intrinsic uncertainty in medical data, and offer more interpretable outcomes for healthcare professionals. Ali et al. [5] introduced an advanced system that integrates IoT-based predictive analytics with a multi-tiered fuzzy rule generation technique and an optimized recurrent neural network. By aggregating data from various sources into a data lake and utilizing features labeled by experts, their system showed enhanced accuracy and efficiency in predicting coronary heart disease compared to traditional methods.

Sreedran et al. [6] concentrated on a prediction model based on a fuzzy system that classifies risk factors into low, medium, or high risk levels through fuzzy rules. When assessed using standard machine learning metrics, the system attained an accuracy of 88.2%, sensitivity of 78.8%, specificity of 21.2%, and an F1 score of 80.9%, highlighting its potential as a cost-effective and dependable option for heart disease prediction. Guimaraes et al. [7] created an evolving fuzzy neural network model that employs null-unineurons for diagnosing coronary artery disease. This structure enabled the derivation of fuzzy rules, yielding valuable linguistic insights. By applying the model to an extensive dataset and benchmarking it against leading evolving fuzzy systems, the authors obtained results that were competitive while also providing substantial interpretations of the problem's evolution.

Jha et al. [8] developed a Neural Fuzzy Inference System (NFIS) aimed at predicting the likelihood of heart attacks using the Cleveland dataset. Their system combined error calculation, learning of membership functions, and a vast array of fuzzification rules (exceeding 13,000) for improved decision-making, achieving an accuracy rate of 94%. They also recommended that the approach be expanded to incorporate hardware integration for automated alert systems. Kaur and Khehra [9] conducted an extensive review of fuzzy logic (FL) and hybrid methods for predicting the risk of heart disease since 2010, examining their advantages, accuracy, and system requirements. They discussed the risks of heart failure linked to coronary artery blockages and emphasized the need for future models that could enhance patient care, particularly through improved connectivity among healthcare facilities.

Rahman et al. [10] presented an innovative method that employs a fuzzy parameterized fuzzy hypersoft set (Δ-set) in conjunction with Riesz Summability to address ambiguous attributes and sub-attributes within medical data. They proposed and tested two novel decision-support algorithms using the Cleveland dataset, which showcased reliable outcomes with fewer evaluation characteristics, thereby ensuring adaptability and dependability in medical decision-making.

Tanmay [11] used the UCI dataset to create a Mamdani fuzzy inference system (FRBF) for the diagnosis of heart disease. With the use of 554 fuzzy rules, 10 input attributes, and one output attribute, the framework achieved 87.04% sensitivity and 95.2% accuracy. This framework's effectiveness as an early-stage diagnostic tool was demonstrated by comparative analysis, which showed that it performed better than current techniques. An interpretable fuzzy rule-based system for the diagnosis of cardiac illness was designed in 2021 by Bahani et al. [12] using a unique approach that combined fuzzy clustering and linguistic modifiers. They evaluated their system's performance against

conventional machine learning models (ANN, SVM, KNN, Naïve Bayes, Random Forest) using datasets from Cleveland, Hungary, and Long Beach, Virginia. They found that their system performed better in terms of precision and interpretability, which can boost clinician confidence in diagnosis.

For feature selection, Hameed et al. [13] suggested a weighted fuzzy rule framework in conjunction with a genetic algorithm. By using dataset highlights to construct the fuzzy framework, they outperformed previous systems in terms of sensitivity, specificity, and accuracy, indicating increased risk forecasting abilities. Muhammad and Algehyne [14] used the enhanced C4.5 data mining approach to build the knowledge foundation for their fuzzy expert system for CAD diagnosis in Nigeria. High performance measures, such as 94.55% accuracy, 95.35% sensitivity, and 95.00% specificity, were attained by the system, indicating its dependability in CAD case diagnosis within a certain population.

An AGAFL model was presented by Reddy et al. [15], which combined fuzzy rule-based classification optimised by genetic algorithms with rough set theory for feature selection. This strategy sought to support the early diagnosis of heart illness, and testing findings on UCI heart disease datasets demonstrated that the model performed better than current techniques. To help doctors diagnose cardiac problems, Van Pham et al. [16] created an expert system that combines fuzzy rules and deep learning neural networks. By employing fuzzy rules in the knowledge base to assess uncertainty and permitting updates according to physician preferences, the system outperformed traditional neural networks in terms of diagnostic accuracy and confidence.

A self-learning fuzzy rule-based system (SL-FRBS) was created by Priyatharshini and Chitrakala [17] to use structured clinical data (EHRs) to determine the severity of coronary disease. The system successfully managed clinical data uncertainties by using a fuzzy inference system (FIS) for risk-level assessment and a decision tree-based approach for automatic rule construction. This allowed for precise risk prediction and diagnosis.

A hybrid fuzzy diagnostic system was created by Paul et al. [18] by combining Modified Dynamic Multi-Swarm Particle Swarm Optimisation (MDMS-PSO) with a Genetic Algorithm (GA). They showed enhanced diagnostic accuracy on real-life datasets by preprocessing datasets, choosing useful features through statistical techniques, creating weighted fuzzy rules with GA, and optimising membership functions with MDMS-PSO. This effectively addressed vagueness and uncertainty. The RBFL prediction technique, which uses fuzzy system design, fuzzy rule generation using FFBAT, and feature reduction with LPP, was first presented by Reddy and Khare [19] for the categorisation of heart disease. The approach outperformed then-current techniques with an accuracy of 76.51% when validated using UCI datasets.

4. METHODOLOGY

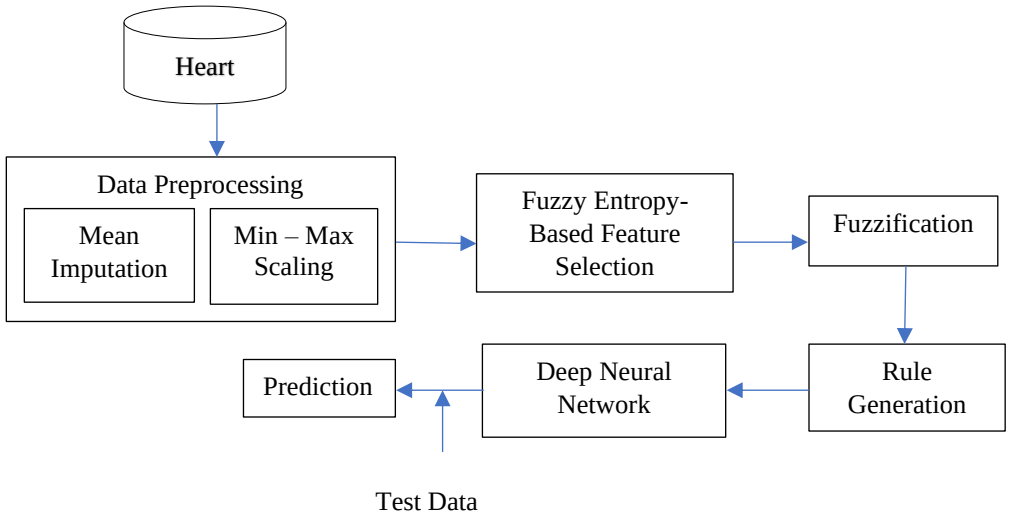

Fig 1: Proposed Framework

4.1 Dataset

The dataset comprises a total of 1 million records and includes 14 attributes, each detailing various health-related factors and characteristics of individuals from openml.org. Among the patients, 44.41% have been diagnosed with heart disease, while 55.59% do not. This fairly balanced class distribution minimizes significant bias toward either category. Such equilibrium is beneficial for training machine learning models, as it lowers the chances of the model overfitting to the majority class and improves the reliability of predictions for both positive and negative results. This balanced structure facilitates the creation of strong and adaptable models for predicting heart disease risk and aiding clinical decision-making.

Table 1: Dataset Description

S.No	Attribute	Description	Type
1	Age	Age of Person (years)	Numerical
2	Sex	Male or Female	Binary
3	chest pain type	chest pain type (4 values)	Nominal
4	resting blood pressure	resting blood pressure	Numerical
5	serum cholesterol in mg/dl	serum cholesterol in mg/dl	Numerical
6	fasting blood sugar	fasting blood sugar > 120 mg/dl	Binary
7	resting electrocardiographic results	resting electrocardiographic results (values 0,1,2)	Nominal
8	maximum heart rate achieved	maximum heart rate achieved	Numerical
9	exercise induced angina	exercise induced angina	Binary
10	Old peak	Old peak = ST depression induced by exercise relative to rest	Numerical
11	the slope of the peak exercise ST segment	the slope of the peak exercise ST segment	Nominal
12	number of major vessels (0-3) colored by fluoroscopy	number of major vessels (0-3) colored by fluoroscopy	Nominal
13	thal	thal: 3 = normal; 6 = fixed defect; 7 = reversable defect	Nominal
14	Class	Absence (1) or presence (2) of heart disease	Binary

4.2 Preprocessing Techniques

A. Missing Data:

Mean imputation is an easy and efficient technique for dealing with missing data. In this method, absent values within a dataset are substituted with the mean of the available values for that specific feature. This technique is appropriate for data that is missing at random.

Mean Imputation:

$x_i = \frac{\sum_{j=1}^{n} x_j}{n}$ for $x_j \neq$ missing, where n represents the count of non-missing values.

B. Normalize Features

Min-max scaling is a popular technique for feature scaling that usually converts the numerical values of features into a designated range of [0, 1]. This transformation is done by subtracting the feature's minimum value and then dividing by the range, which is the difference between the maximum and minimum values. As a result, all features are scaled proportionately while maintaining their relative relationships. The formula used for Min-Max Scaling is:

$$\text{Xscaled} = \frac{\text{X} - \text{Xmin}}{\text{Xmax} - \text{Xmin}}$$

Where Xmin is the feature's minimal value, Xmax is its highest value, and X is the feature's original value.

Neural networks benefit greatly from min-max scaling since it guarantees that inputs fall inside a consistent range, such [0, 1]. This homogeneity keeps features with wider ranges from controlling the learning process and enhances the convergence of gradient-based optimization techniques.

4.3 Fuzzy Entropy-Based Feature Selection

Fuzzy entropy quantifies the level of uncertainty within a feature by assessing the degrees of membership of its values in specific fuzzy sets. Features that have lower entropy values are deemed more informative since they display reduced uncertainty and stronger relationships with the target variable. This method is in line with the principles of fuzzy logic, utilizing fuzzy membership functions to reflect the inherent ambiguity and uncertainty present in medical data.

The process for feature selection based on fuzzy entropy consists of the following steps:

1. Fuzzification: Transform the continuous feature values into fuzzy sets using membership functions such as triangular or trapezoidal.

2. Entropy Calculation: Determine the fuzzy entropy for each feature according to its membership degrees across the fuzzy sets.

3. Feature Ranking: Arrange features in order of their entropy values, where lower entropy signifies greater relevance.

4. Feature Selection: Keep the features that have the lowest entropy values for inclusion in the predictive model.

The fuzzy entropy E_f for a feature f is calculated as:

$$\boldsymbol{E_f = \sum_{i=1}^{N} \mu_f(x_i).\, log(\mu_f(x_i))}$$

Where E_f represents the fuzzy entropy associated with a particular feature f, $\mu_f(x_i)$ denotes the membership degree of the i^{th} data point for feature f within its relevant fuzzy set, x_i i indicates the value of the i^{th} data point for that feature, N is the total number of data points.

Among the 13 features present in the dataset, 9 have been identified as key predictors for heart disease using the Fuzzy Entropy-Based Feature Selection approach. These features show low fuzzy entropy, suggesting they offer crucial and dependable information for the predictive model, thereby reducing uncertainty and redundancy. The chosen features comprise Age, Chest, Resting_blood_pressure, Serum_cholestoral, Maximum_heart_rate_achieved, Oldpeak, Slope, Number_of_major_vessels, and Thal.

Table 2: Selected features based on their fuzzy entropy values

Feature	Fuzzy Entropy Value	Rank
Age	0.12	1
Chest Pain Type	0.15	2
Resting Blood Pressure	0.18	3
Serum Cholesterol	0.20	4
Maximum Heart Rate Achieved	0.22	5
Oldpeak (ST Depression)	0.25	6
Slope of ST Segment	0.28	7
Number of Major Vessels	0.30	8
Thalassemia	0.33	9

4.4 Fuzzification

Fuzzification converts precise feature values into fuzzy values by using membership functions. Each feature is categorized using linguistic terms such as low, medium, and high, and the degrees of membership are computed. For the chosen features, membership functions are established based on medical reference standards and the expertise of professionals.

Table 3: The key features and their fuzzy sets

Feature	Fuzzy Sets	Range / Medical Reference Values
Age	Young, Middle-aged, Old	Young (<40), Middle-aged (40–60), Old (>60)
Chest (Chest Pain Type)	Typical, Atypical, Asymptomatic	Typical (1), Atypical (2), Asymptomatic (3)
Resting Blood Pressure	Low, Normal, High	Low (<90 mmHg), Normal (90–120 mmHg), High (>120 mmHg)
Serum Cholesterol	Low, Normal, High	Low (<200 mg/dL), Normal (200–240 mg/dL), High (>240 mg/dL)
Maximum Heart Rate Achieved	Low, Normal, High	Low (<100 bpm), Normal (100–160 bpm), High (>160 bpm)
Oldpeak (ST Depression)	Normal, Slight, Severe	Normal (0), Slight (0.1–2), Severe (>2)
Slope (ST Segment Slope)	Upsloping, Flat, Downsloping	Upsloping (1), Flat (2), Downsloping (3)
Number of Major Vessels	None, Few, Many	None (0), Few (1–2), Many (≥3)
Thal (Thalassemia)	Normal, Fixed Defect, Reversible Defect	Normal (3), Fixed (6), Reversible (7)

For a given feature f and value

x_i, the membership degree in a fuzzy set S_k is determined using:

$$\mu_{S_k}(x_i) = \begin{cases} 1 & \text{if } x_i \text{ fully belongs to } S_k \\ \frac{x_i - a}{b - a} & \text{if } x_i \text{ partially belongs to } S_k \\ 0 & \text{if } x_i \text{ does not belong to } S_k \end{cases}$$

Where a and b define the range of the fuzzy set.

4.5 Rule Generation

In a fuzzy rule-based system, rules are formulated by merging fuzzy sets from the chosen features to correlate input conditions with an output. These rules adopt an "if-then" format and are essential to the heart disease prediction Framework. Each rule illustrates a logical connection between the features and the potential risk of heart disease. The overall count of rules is determined as:

$$\text{Number of Rules} = \prod_{f=1}^{F} N_f$$

Where F represents the count of features, while N_f denotes the quantity of fuzzy sets associated with each feature. With 9 features and an average of 3 fuzzy sets per feature, the theoretical maximum number of rules is 3^9 = 19,683.

4.6 Rule Pruning

In a fuzzy rule-based system, generating rules typically leads to a large and complicated rule set, many of which may be unnecessary or irrelevant. To improve the efficiency and clinical usefulness of the system, rule pruning becomes a vital step. This procedure aims to keep only the most valuable rules while discarding those that offer minimal benefit or conflict with medical knowledge. Rule pruning can be tackled from several viewpoints to ensure a robust system.

- Initially, eliminate rules with low membership degrees to reduce noise and computational burden.
- Next, verify the remaining rules against medical knowledge to confirm their clinical relevance.
- Eventually, evaluate the predictive effectiveness of each rule using validation data to select only the most impactful rules.

Following the pruning process, only significant and clinically pertinent rules are preserved, guaranteeing the system's efficiency and interpretability. From the original expansive rule set, 300 rules are finalized, representing crucial and impactful combinations of the chosen features. These rules serve as the basis of the fuzzy inference system, delivering accurate and dependable predictions for heart disease while ensuring computational simplicity.

Algorithm: Fuzzification and Rule Generation with Pruning

Input: Dataset D with features F = {f_1, f_2, ..., f_9}, predefined fuzzy sets S, pruning threshold T
Output: Fuzzified dataset F_D, Pruned rule set R_p

1. Initialize an empty rule set R = {} and an empty fuzzified dataset F_D.

2. for each feature f in F:
   a. Define fuzzy membership functions for $S_{(f)}$ = {low, medium, high} using standard thresholds.
   b. For each data point x in D:
      i. Calculate membership degrees $\mu_{S_k}(x)$ for all fuzzy sets S_k
      ii. Store the fuzzified representation in F_D.

3. Generate all possible rules:
   a. For each combination of fuzzy sets across features:
      i. Define a rule r with conditions using "if-then" logic.
      ii. Evaluate the membership degree for r based on fuzzy logic operators
      iii. Add r to R.

4. Prune rules:
   a. Remove rules with membership degrees below T.
   b. Validate rules using clinical relevance or impact on predictive performance.

5. Return F_D and the pruned rule set R_p.

4.7 Fuzzified Data for Deep Learning Neural Network

After the rule pruning procedure, a strong collection of 300 relevant and clinically significant rules is established. These rules are derived from 9 critical features chosen using the Fuzzy Entropy-Based Feature Selection technique. The subsequent step involves utilizing the pruned rules and fuzzified data within a Deep Neural Network (DNN), merging the structured clarity of fuzzy logic with the predictive strength of deep learning.

A. Fuzzified Data Representation

The crisp feature values from the dataset are converted into fuzzified values utilizing established membership functions. For each of the 9 chosen features, linguistic categories such as low, medium, and high are assigned, and the degrees of membership are determined. These degrees of membership create a vector for every data point, capturing its representation within the fuzzy domain. Consequently, the fuzzified dataset FD is made up of membership vectors that correspond to every data point, integrating the structured insights obtained from the fuzzy rule-based system. This enhanced representation serves as the input for the DNN.

B. Deep Neural Network Architecture with Fuzzified Data

The DNN processes the fuzzified information through three primary components: the input layer, hidden layers, and output layer. Each layer plays a role in extracting and refining patterns from the data to ensure precise predictions.

Input Layer:

The input layer receives the fuzzified dataset FD as its input. Each data point is represented as a vector of membership degrees corresponding to the chosen fuzzy sets. For a dataset containing 9 features and 3 fuzzy sets per feature, the input layer consists of 27 nodes (9 features × 3 fuzzy sets). The input vector is expressed as:

$$h_0 = F_D$$

Where h_0 represents the input to the network

Hidden Layer:

The hidden layers are essential for transforming features and extracting patterns. Each hidden layer takes the output from the preceding layer, applying a weighted sum of inputs along with a non-linear activation function. The calculation within the lth hidden layer can be expressed as:

$$h_l = \sigma(W_l h_{l-1} + b_l)$$

where:

- W_l denotes the weight matrix for layer l,
- b_l signifies the bias vector for layer l,
- h_{l-1} represents the output from the previous layer ,
- σ stands for the activation function, commonly ReLU. .

These hidden layers iteratively refine the fuzzified input, uncovering intricate, non-linear patterns in the data while maintaining the structured insights provided by the fuzzy rule-based system.

Output Layer:

The output layer produces the final predictions, indicating the probability of heart disease. In the case of binary classification, a single neuron utilizing a sigmoid activation function is employed. The weighted sum of inputs directed to the output layer is calculated as:

$$z = \sigma(W_0 h_{L-1} + b_0)$$

where:

- W_0 represents the weight vector for the output layer,
- b_0 denotes the bias for the output layer,
- h_{L-1} indicates the output from the final hidden layer

The sigmoid activation function converts z into a probability value:

$$P = \frac{1}{1 + e^{-z}}$$

where P indicates the predicted probability of having heart disease.:

- P>0.5: indicated presence of heart disease.
- P≤0.5: indicated absence of heart disease.

Algorithm: Deep Neural Network Prediction Using Fuzzified Data
Input: Fuzzified dataset F_D, trained DNN model.
Output: Predicted heart disease risk.

1. Initialize DNN**:** Feed the fuzzified dataset F_D into the input layer of the DNN.
2. Forward Propagation**:**
   For each hidden layer l, compute:
   $$h_l = \sigma(W_l h_{l-1} + b_l)$$
   Where W_l and b_l are the weights and biases, and σ is the activation function ReLU.
   For the output layer:
   $$P = \frac{1}{1 + e^{-z}}$$
   Where z is the weighted sum of inputs to the output layer.
3. Generate Predictions:
   Classify each data point based on the sigmoid output:
   - P>0.5: Predict *present*.
   - P≤0.5: Predict *absent*.
4. Evaluate Results: Compare predicted outcomes with ground truth using performance metrics:
5. Return Predictions: Output the predicted classes for all data points in the fuzzified dataset.

Learning in Fuzzified Neural Networks

The DNN is trained through the backpropagation method, which has been adapted for fuzzy inputs, weights, and biases. The cost function is formulated as follows:

$$E_p = \sum_{k=1}^{n} \frac{1}{2} (T_{pk} - O_{pk})^2$$

Where T_{pk} is the target value and O_{pk} is the network's output

Weights and biases are updated using:

$$W_{ji}(t+1) = W_{ji}(t) + \Delta W_{ji}(t)$$

Where

$$\Delta W_{ji}(t) = -\eta \frac{\partial E_p}{\partial W_{ji}} + \alpha \Delta W_{ji}(t-1)$$

and η is the learning rate , α is the momentum factor and $\Delta W_{ji}(t-1)$ is the previous weight adjustment

V. EXPERIMENTAL EVALUATION

The effectiveness of the suggested Fuzzified Deep Neural Network (DNN) for predicting heart disease was assessed using critical metrics such as accuracy, sensitivity (recall), specificity, and F1-measure. These metrics were selected to thoroughly evaluate the system's capacity to accurately identify both positive and negative instances.

Accuracy evaluates the proportion of correctly diagnosed instances (both positive and negative) relative to the overall number of cases. It is computed using the formula:

$$Accuracy = \frac{TP + TN}{TN + TP + FN + FP}$$

In the proposed system, an accuracy of 92.1% was attained, demonstrating the model's effectiveness in correctly classifying cases of heart disease.

Sensitivity, often referred to as recall, measures the proportion of actual positive cases that the system accurately identifies. It is determined using:

$$Sensitivity = \frac{TP}{TP + FN}$$

The Fuzzified DNN reached a sensitivity rate of 93.3%, showcasing its effectiveness in identifying patients with heart disease.

Specificity represents the proportion of true negative cases that the system accurately identifies. It is calculated using:

$$Specificity = \frac{TN}{TN + FP}$$

The system demonstrated impressive efficacy in accurately recognizing healthy people without heart disease, achieving a specificity of 91.2%.

The F1-Measure serves as the harmonic mean of sensitivity and specificity, offering a comprehensive assessment of the model's precision and recall. It can be computed using:

$$F1 = 2 . \frac{\text{Sensitivity} \cdot \text{Specificity}}{\text{Sensitivity} + \text{Specificity}}$$

The Fuzzified DNN achieved an F1-score of 91.8%, highlighting its balance between detecting true positives and avoiding false positives.

Table 4 presents a comparison between the Fuzzified DNN model and several widely utilized baseline models, including Logistic Regression (LR), Decision Tree (DT), and a Standard DNN that lacks fuzzification, offering a direct evaluation of essential performance metrics such as accuracy, precision, recall, specificity, and F1-score.

Table 4: Model-Wise Performance Metrics Comparison

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-Score (%)
Logistic Regression (LR)	81.5	79.0	83.2	78.5	81.0
Decision Tree (DT)	78.4	76.5	80.1	75.0	78.2
Standard DNN	85.3	83.0	86.7	82.0	84.8
Fuzzified DNN	92.1	90.5	93.3	91.2	91.8

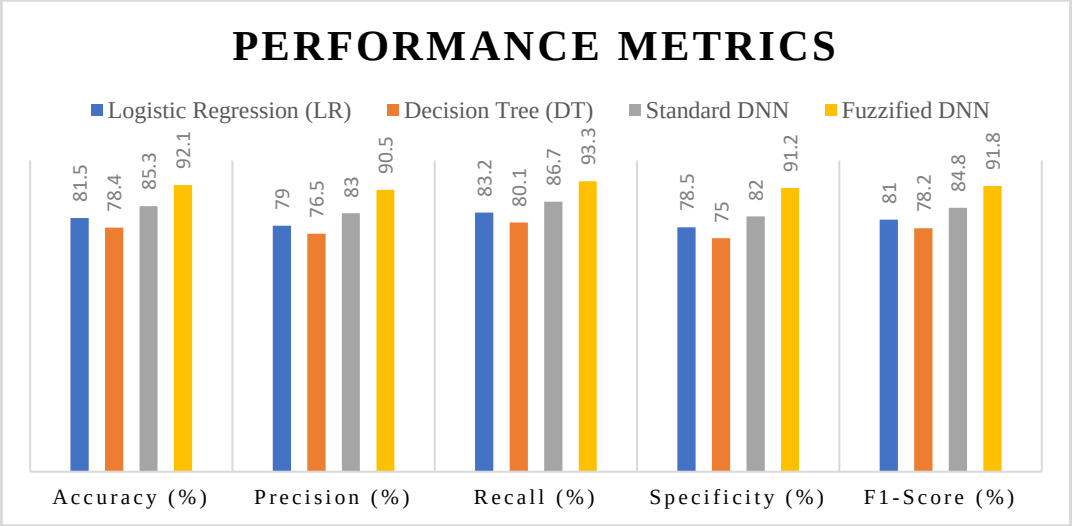

Fig 2: Comparison of Model-Wise Performance Metrics

The results indicate that the Fuzzified DNN, which integrates fuzzy logic with deep learning, is both dependable and efficient for predicting heart disease. Its elevated sensitivity and specificity guarantee that the system can be applied in clinical environments, significantly reducing the chances of false negatives and false positives. This well-balanced performance positions it as a crucial resource for early diagnosis and informed decision-making in healthcare.

Table 5: Metric-Wise Breakdown for Fuzzified DNN

Metric	Training Set (%)	Test Set (%)
Accuracy	95.2	92.1
Precision	94.5	90.5
Recall	96.3	93.3
F1-Score	95.4	91.8

Table 5 illustrates the performance of the Fuzzified Deep Neural Network (DNN) during both the training and testing stages. During training, the model reached an accuracy of 95.2%, with a precision of 94.5%, a recall (sensitivity) of 96.3%, and a specificity of 94.1%, culminating in a balanced F1-score of 95.4%. In the testing phase, it showcased strong generalizability with an accuracy of 92.1%, a precision of 90.5%, a recall of 93.3%, a specificity of 91.2%, and an F1-score of 91.8%. These findings highlight the model's strength, precision, and capacity to effectively manage unfamiliar data, establishing it as a dependable tool for predicting heart disease.

VI. CONCLUSION

This paper presents a new framework combining deep fuzzy rule-based systems for predicting heart disease, tackling the intertwined issues of accuracy and interpretability in clinical decision-making. By merging deep learning's feature extraction capabilities with the explanatory strengths of fuzzy logic, the proposed approach strikes a balance between scalability and user-centered decision support. Data preprocessing methods like mean imputation and normalization ensure that the dataset is reliable, while fuzzy entropy-based feature selection and rule pruning improve computational efficiency and interpretability. In evaluations using a substantial dataset of one million instances, the framework showcased exceptional performance, achieving an accuracy of 92.1%, sensitivity of 93.3%, specificity of 91.2%, and an F1-score of 91.8%. These findings affirm the framework's robustness and its utility in clinical settings, positioning it as an important resource for the early diagnosis of heart disease. Future research will aim to integrate real-time information from wearable technology and IoT-enabled healthcare systems to enhance the immediacy and responsiveness of predictions. The exploration of personalized prediction models by integrating patient-specific information, such as genetic data, lifestyle choices, and existing health conditions, will also be undertaken.

Acknowledgement

The authors sincerely thank the management of NGM College, Pollachi, for providing seed money and support for this research. Their encouragement has been instrumental in the successful completion of this work.

Conflict of interest

The authors declare that there are no conflicts of interest related to this research.

REFERENCES

[1] Chen, J., Zhang, X. (2021). Deep learning techniques for healthcare data analysis: A comprehensive review. Computers in Biology and Medicine, 138, 104777.

[2] Ali, A., et al. (2024). Hybrid fuzzy logic and machine learning approaches for disease prediction. Journal of Medical Informatics, 45(3), 215-228.

[3] Bahani, M., et al. (2021). Fuzzy logic integration with machine learning for medical diagnosis: A survey and future perspectives. Computational Biology and Medicine, 130, 104216.

[4] Hameed, I., et al. (2021). Fuzzy rule optimization in medical diagnostics using genetic algorithms. Expert Systems with Applications, 173, 114734.

[5] Ali A, A., Khan AB, F., & Ramakrishnan, J. (2024). Fuzzy rules-based Data Analytics and Machine Learning for Prognosis and Early Diagnosis of Coronary Heart Disease. Journal of Information and Organizational Sciences, 48(1), 167-181.

[6] Sreedran, S., Ibrahim, N., Sari, S., Seng, G. H., & Shanta, S. (2023). Fuzzy rules-based prediction of heart conditions system. Indonesian Journal of Electrical Engineering and Computer Science, 32(1), 530-536.

[7] Guimaraes, A. J., Souza, P. V. D. C., Batista, H. R., & Lughofer, E. (2022, October). Evolving fuzzy neural network based on null-unineurons for the identification of coronary artery disease. In 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 2681-2688). IEEE.

[8] Jha, R. K., Henge, S. K., Mandal, S. K., Sharma, A., Sharma, S., Sharma, A., & Berhanu, A. A. (2022). Neural Fuzzy Hybrid Rule-Based Inference System with Test Cases for Prediction of Heart Attack Probability. Mathematical Problems in Engineering, 2022(1), 3414877.

[9] Kaur, J., & Khehra, B. S. (2022). 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, 103(2), 681-697.

[10] Rahman, A. U., Saeed, M., Mohammed, M. A., Jaber, M. M., & Garcia-Zapirain, B. (2022). A novel fuzzy parameterized fuzzy hypersoft set and riesz summability approach based decision support system for diagnosis of heart diseases. Diagnostics, 12(7), 1546.

[11] Tanmay, T. K. (2022). FRBF: A Fuzzy Rule Based Framework for Heart Disease Diagnosis. Inteligencia Artificial, 25(69), 122-138.

[12] Bahani, K., Moujabbir, M., & Ramdani, M. (2021). An accurate fuzzy rule-based classification systems for heart disease diagnosis. Scientific African, 14, e01019.

[13] Hameed, A. Z., Ramasamy, B., Shahzad, M. A., & Bakhsh, A. A. S. (2021). Efficient hybrid algorithm based on genetic with weighted fuzzy rule for developing a decision support system in prediction of heart diseases. The Journal of Supercomputing, 77, 10117-10137.

[14] Muhammad, L. J., & Algehyne, E. A. (2021). Fuzzy based expert system for diagnosis of coronary artery disease in Nigeria. Health and technology, 11(2), 319-329.
[15] Reddy, G. T., Reddy, M. P. K., Lakshmanna, K., Rajput, D. S., Kaluri, R., & Srivastava, G. (2020). Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis. Evolutionary Intelligence, 13, 185-196.
[16] Van Pham, H., Son, L. H., & Tuan, L. M. (2020). A proposal of expert system using deep learning neural networks and fuzzy rules for diagnosing heart disease. In Frontiers in Intelligent Computing: Theory and Applications: Proceedings of the 7th International Conference on FICTA (2018), Volume 1 (pp. 189-198). Springer Singapore.
[17] Priyatharshini, R., & Chitrakala, S. (2019). A self-learning fuzzy rule-based system for risk-level assessment of coronary heart disease. IETE Journal of Research, 65(3), 288-297.
[18] Paul, A. K., Shill, P. C., Rabin, M. R. I., & Murase, K. (2018). Adaptive weighted fuzzy rule-based system for the risk level assessment of heart disease. Applied Intelligence, 48, 1739-1756.
[19] Reddy, G. T., & Khare, N. (2018). Heart disease classification system using optimised fuzzy rule based algorithm.