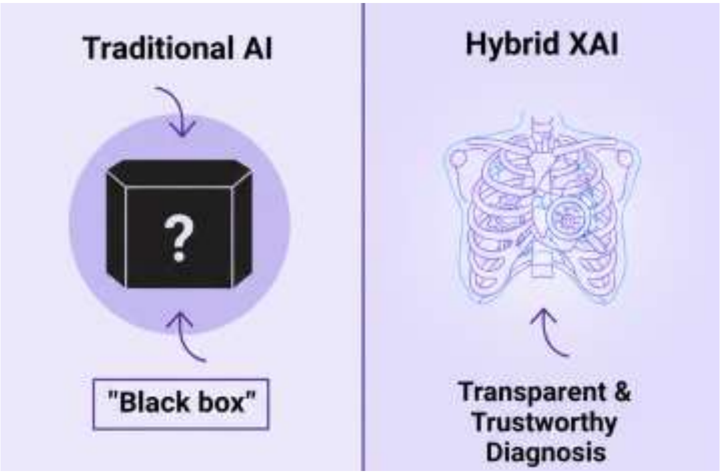
Fusing Neural Networks and Fuzzy Logic for Interpretable and Accurate Medical Diagnosis: A Hybrid Explainable AI (XAI) Framework



Supervisor Name:
Dr. Atul Kumar Verma

Name: Sanya Goyal
23FE10CSE00684

Name: Tanisha Verma
23FE10CSE00095

Section: B

Problem Statement and Definition

Statement: How can a hybrid model combining a Neural Network for disease classification and a Fuzzy Logic system for rule-based explanation provide transparent and interpretable insights into a medical diagnosis?

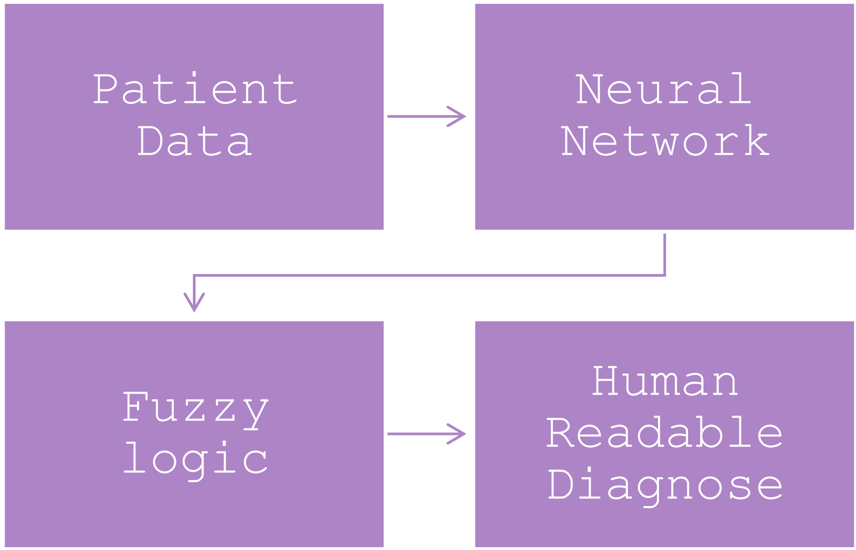
The research will focus on tackling the lack of interpretability in AI models for healthcare as well as the trade-offs issue between interpretability and accuracy in existing AI models.

Application of AI approaches

- **Machine Learning (ML) in Medical Diagnosis:** Used for predictive modeling and disease risk estimation.
- **Deep Learning / Neural Networks:** Extract complex features automatically.
- **Fuzzy Logic Systems:** Handle uncertainty and vagueness in medical data and [provide human-readable rules.
- **Hybrid AI (Proposed XAI Approach):**

➢ Combines Neural Networks (for accuracy) with Fuzzy Logic (for interpretability).

➢ Converts NN outputs into linguistic rules and confidence



Reason to Opt AI Approach

- **Complexity of Medical Data:** Medical data are complex, high-dimensional, and multimodal leading to occasional missed hidden patterns by the traditional methods. AI captures these hidden patterns efficiently.
- **Accuracy and Efficiency:** Neural Networks ensure high diagnostic accuracy and automation of analysis saves the doctor time.
- **Handling Uncertainty:** Fuzzy Logic manages vague or uncertain medical inputs.
- **Explainability Requirement:** Clinicians need transparent, trustworthy models. Hybrid AI (NN + FL) provides explainable predictions and are used here.

Implementation: Methodology

A hybrid CNN-Fuzzy Inference System (FIS) model is developed for heart disease diagnosis. The CNN learns complex clinical patterns and outputs probability scores, which the FIS translates into interpretable, rule-based risk levels. This cascaded approach combines deep learning accuracy with fuzzy logic explainability for transparent medical predictions.

Collab Link:
https://colab.research.google.com/drive/1lsEGXlD5ylH9PPl4uCV2urqnff4ZOUSE?usp=sharing

Implementation: Workflow

Step 1 – Data Acquisition

- Dataset: UCI Cleveland Heart Disease (303 records, 14 features).
- Missing values imputed with median.

Step 2 – Preprocessing & Feature Engineering

- Created derived features: hr_reserve, bp_category, chol_category, cp_severity, age_group.
- Neural Network uses 15 numerical features.
- Fuzzy Logic uses 4 key features: trestbps, chol, thalach, oldpeak.
- Data scaled using StandardScaler and split 80/20.

Step 3 – Neural Network (CNN)

- Architecture: Conv1D → MaxPool → Conv1D → Dense layers.
- Learns patterns and outputs heart disease probability.
- Test accuracy ≈ 82%, AUC ≈ 0.88.

Step 4 – Fuzzy Logic (FIS)

- Inputs: CNN probability, cholesterol, thalach, oldpeak.
- Rules use linguistic terms (LOW, MEDIUM, HIGH).
- Produces defuzzified heart risk score (0–1).

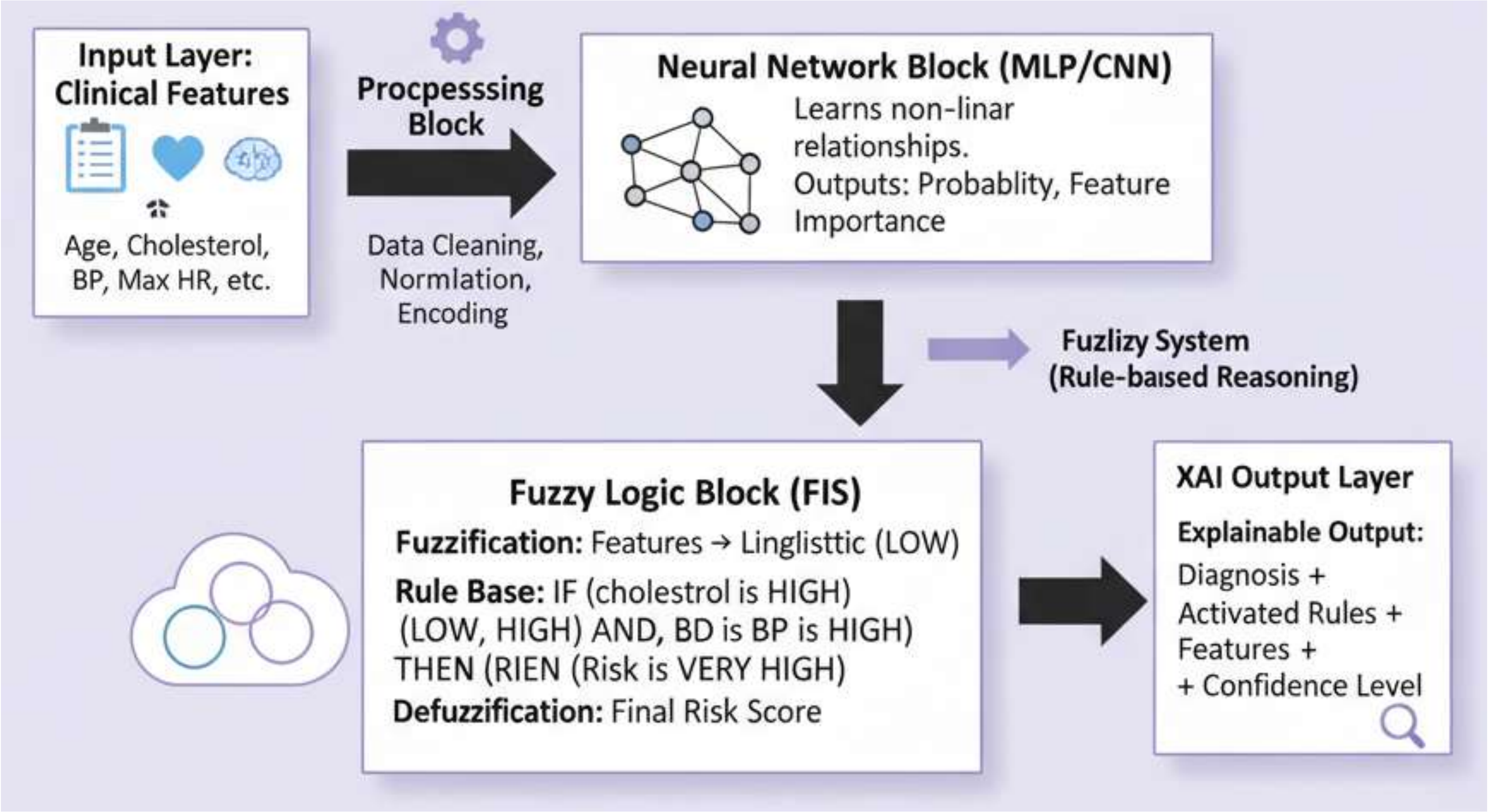
Step 5 – Hybrid Integration

- CNN output + key features → FIS → interpretable heart risk.
- Threshold ≥ 0.5 indicates disease presence.

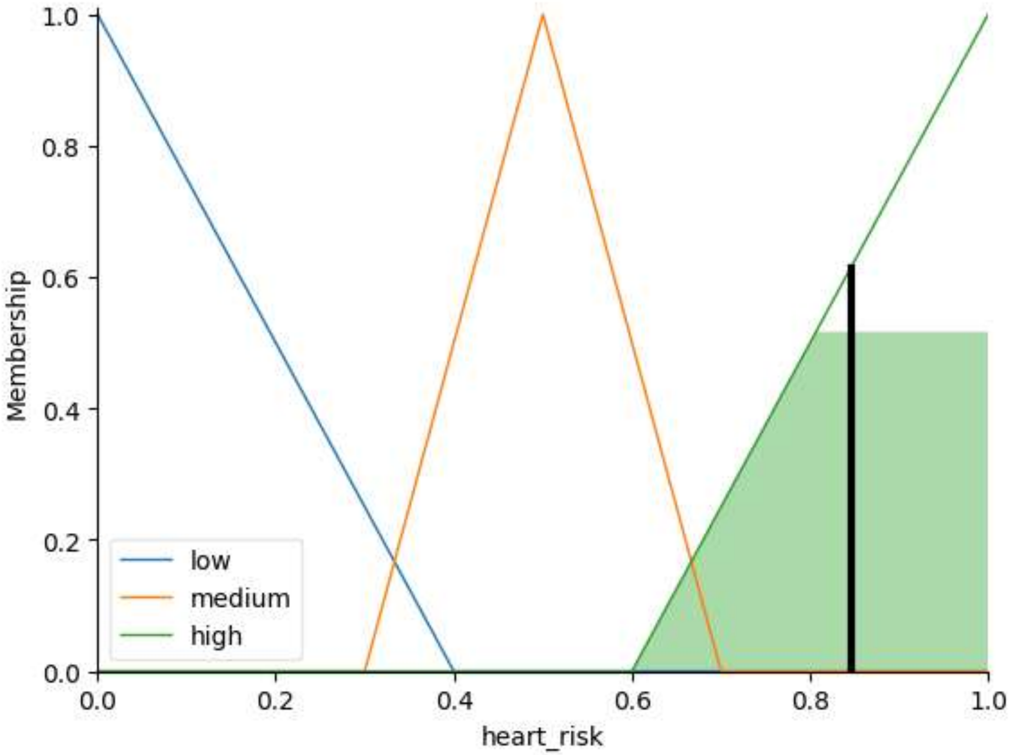
Step 6 – Explainability & Validation

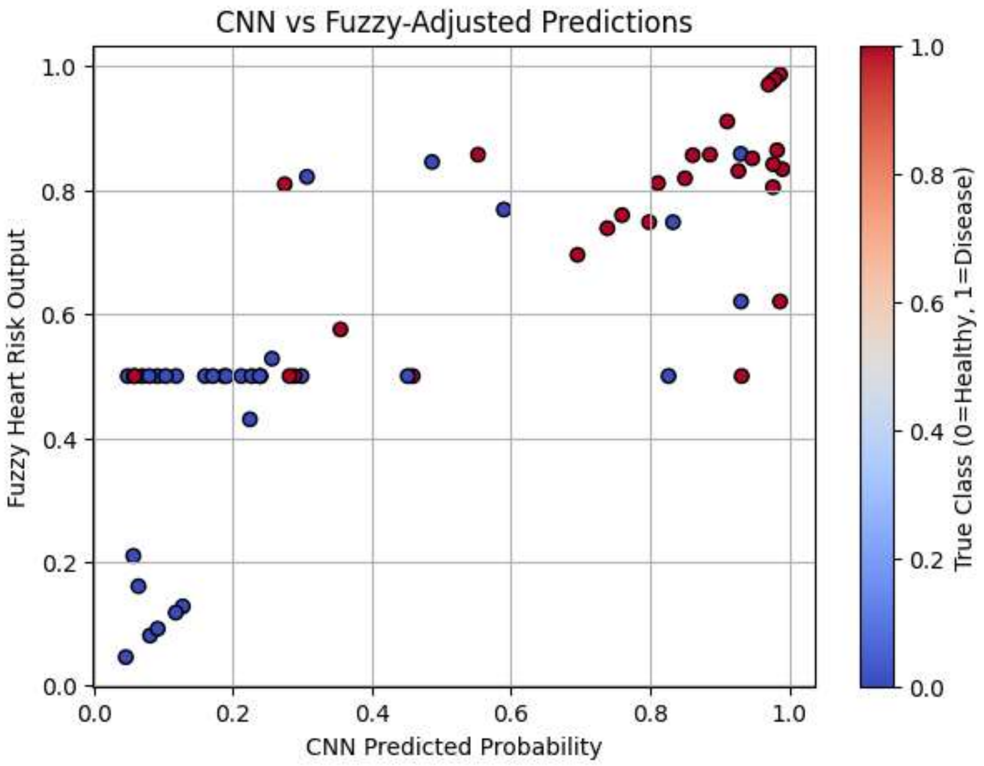
- Generates human-readable explanations and rule activations.
- Hybrid model accuracy ≈ 80%, AUC ≈ 0.86.
- Provides interpretable, confidence-based diagnosis.

System Architecture



Result and Validation





The output of a fuzzy logic system
designed to calculate a "heart_risk" score.

```python
from sklearn.metrics import accuracy_score, roc_auc_score, confusion_matrix

acc = accuracy_score(y_test, fuzzy_preds)
auc = roc_auc_score(y_test, fuzzy_outputs)
cm = confusion_matrix(y_test, fuzzy_preds)

print("Hybrid CNN-FIS Model Evaluation:")
print(f"Accuracy: {acc:.3f}")
print(f"AUC: {auc:.3f}")
print("Confusion Matrix:\n", cm)
```

```
Hybrid CNN-FIS Model Evaluation:
Accuracy: 0.803
AUC: 0.860
Confusion Matrix:
 [[24  9]
 [ 3 25]]
```

Accuracy and AUC of the Hybrid Model

```python
explain_table = pd.DataFrame(rows)
print("\nExplainability Table – Sample Predictions:\n")
print(explain_table.to_string(index=False))
```

```
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 142ms/step
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 183ms/step
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 143ms/step
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 157ms/step
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 71ms/step

Explainability Table – Sample Predictions:

 Index  True Label  CNN Prob  Cholesterol  Thalach  Fuzzy Heart Risk
     0           0     0.486        271.0    182.0             0.846
     5           0     0.081        178.0     96.0             0.846
    10           1     0.861        281.0    103.0             0.857
    15           1     0.696        176.0     90.0             0.857
    20           0     0.590        234.0    161.0             0.769
```

Explainability Table

Paper Draft

- https://drive.google.com/drive/folders/1AKAWaJS4O2UonOkJ_NMvdOfxg3LDRlW3?usp=sharing

Conclusion

The proposed hybrid CNN–Fuzzy Inference System effectively bridges the gap between accuracy and interpretability in medical diagnosis. By combining the predictive strength of neural networks with the transparent reasoning of fuzzy logic, the model delivers reliable heart disease predictions along with human-understandable explanations. This approach enhances clinician trust and demonstrates the potential of Explainable AI (XAI) frameworks in developing transparent, and trustworthy healthcare systems.

References

[1] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. Setio, F. Ciompi, M. Ghafoo-rian, J. A. van der Laak, B. van Ginneken, and C. I. S´anchez, “A survey on deep learning in medical image analysis,” Medical image analysis, vol. 42, pp. 60–88, 2017.

[2] A. Rajkomar, E. Oren, K. Chen, A. M. Dai, N. Hajaj, M. Hardt, P. J. Liu, X. Liu, J. Marcus, M. Sun et al., “Scalable and accurate deep learning with electronic health records,” NPJ Digital Medicine, vol. 1, no. 1, p. 18, 2018.

[3] E. J. Topol, “High-performance medicine: the convergence of human and artificial intelligence,” Nature medicine, vol. 25, no. 1, pp. 44–56, 2019.

[4] C. Rudin, “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead,” Nature Machine Intelligence, vol. 1, no. 5, pp. 206–215, 2019.

[5] J. Amann, A. Blasimme, E. Vayena, D. Frey, and V. I. Madai, “Ex-plainability for artificial intelligence in healthcare: a multidisciplinary perspective,” BMC Medical Informatics and Decision Making, vol. 20, no. 1, pp. 1–9, 2020.

[6] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, and N. Elhadad, “Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission,” pp. 1721–1730, 2015.

[7] A. Esteva, K. Chou, S. Yeung, N. Naik, A. Madani, A. Mottaghi, Y. Liu, E. Topol, R. Socher, and J. Dean, “Deep learning-enabled medical computer vision,” NPJ Digital Medicine, vol. 4, no. 1, p. 5, 2021.

[8] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, “Machine learning and data mining methods in diabetes research,” Computational and structural biotechnology journal, vol. 15, pp. 104–116, 2017.

[9] A. Adadi and M. Berrada, “Peeking inside the black-box: A survey on explainable artificial intelligence (xai),” IEEE access, vol. 6, pp. 52 138–52 160, 2018.

Guide Consent Form

AISC Research Paper

Name: SANYA GOYAL 23FE10CSE00684
TANISHA VERMA 23FE10CSE00096
Section: B
Faculty Supervisor: ATUL KUMAR VERMA

PROBLEM STATEMENT:

How can a hybrid model combining a Neural Network for disease classification and a Fuzzy Logic System for rule-based explanation provide transparent and interpretable insights into a medical diagnosis. The research will focus on tackling the lack of interpretability in AI models for healthcare as well as trade-off issues between interpretability and accuracy in existing AI models.

OBJECTIVE:

The main objective is to design, develop and validate a hybrid Explainable AI (XAI) model that achieves high accuracy in medical diagnosis (like skin condition classification) while model aims to deliver reliable, transparent and clinically meaningful results without compromising computational efficiency.

PROBLEM DEFINATION:

Current state of the art Neural Networks (NNs) in medical diagnosis are highly accurate but suffer from a lack of transparency (the "black box" problem), which hinders their adoption by clinicians due to reluctance to trust. The problem is defined as developing a novel hybrid architecture that leverages the NN's predictive power and integrates a Fuzzy logic system to systematically translate the complex, non-linear NN outputs into simpler, verifiable IF-THEN rules, thus providing inherent and post-hoc explanation for the diagnosis.

IDENTIFIED AISC APPROACH:

The identified approach is a hybrid model combining two distinct paradigms

1. Artificial Neural Network (ANN) / Deep learning:
(Used for core pattern recognition & classification task, Eg: classifying skin conditions from images)
2. Fuzzy logic (FL) system:
(used for modelling uncertainty, handling vagueness and rule extraction. It will act as the interpretable layer, translating the NN's probability scores. Eg: IF confidence is HIGH, THEN malignant likelihood is VERY HIGH)

Faculty Supervisor Sign: 8.10.2025.