Fuzzy System in XAI And their Real World Application

Divyanshi Yadav

January 2, 2025

Abstract

Chronic Kidney Disease (CKD) is a severe health condition that affects millions of people worldwide. Early detection of CKD is crucial for effective management and treatment. This paper proposes a novel approach to predict CKD using a hybrid model combining the Adaptive Neuro-Fuzzy Inference System (ANFIS) and an attention mechanism. The model leverages fuzzy rules for handling uncertainty and attention mechanisms for focusing on the most relevant features, aiming to improve the prediction accuracy and interpretability of the model. We evaluate the model on a dataset containing patient information and demonstrate that our hybrid model outperforms traditional machine learning models.

1 Introduction

Chronic Kidney Disease (CKD) is a progressive medical condition characterized by the gradual loss of kidney function, posing a significant public health challenge worldwide. According to global health statistics, CKD affects millions of individuals annually, with the potential to cause severe complications such as cardiovascular disease, kidney failure, and premature mortality. Early diagnosis and intervention are critical to mitigating disease progression, reducing associated healthcare costs, and improving patient outcomes.

Traditional diagnostic approaches for CKD primarily rely on laboratory tests, including serum creatinine levels, glomerular filtration rate (GFR), and proteinuria analysis. While these methods are well-established, they often fall short in detecting early-stage CKD due to their dependency on threshold-based interpretations and the influence of patient-specific factors such as age, gender, and comorbidities. The growing availability of healthcare data and advancements in machine learning (ML) have opened new avenues for leveraging predictive models to enhance early diagnosis and support clinical decision-making. [4]

However, despite the success of ML techniques in improving prediction accuracy, a major limitation persists: the interpretability of complex models. In healthcare, explainability is paramount, as clinicians and healthcare providers must understand and trust the reasoning behind a model's predictions. Black-box models, such as deep learning algorithms, often lack transparency, limiting their adoption in critical domains like CKD diagnosis.

To address these challenges, this paper introduces a novel hybrid framework that combines the Adaptive Neuro-Fuzzy Inference System (ANFIS) with an attention mechanism for CKD prediction. ANFIS integrates the strengths of fuzzy logic and neural networks, offering a powerful approach to handle uncertain and imprecise data while ensuring interpretability. The addition of an attention mechanism further enhances the model's ability to focus on the most relevant features, improving both predictive performance and explainability. By bridging the gap between accuracy and interpretability, the proposed model empowers healthcare professionals with actionable insights while maintaining trust in the system's outputs.

This research makes the following contributions:

We propose a hybrid ANFIS and attention-based model tailored for CKD prediction, leveraging fuzzy logic for uncertainty handling and an attention mechanism for enhanced interpretability. We evaluate the model's performance using a publicly available CKD dataset and compare it against traditional machine learning algorithms. We integrate explainability techniques, including Local Interpretable Model-agnostic Explanations (LIME) , to provide transparency and insights into the model's decision-making process.

2 Related Work

The paper presents a fuzzy logic system for identifying chronic kidney disease (CKD) using eight input parameters, including age and blood pressure. The results show promising accuracy in determining CKD stages. However, limitations include reliance on specific data sets and potential variability in patient conditions. [5]

The paper presents a model for predicting chronic kidney disease (CKD) using an expert system that integrates fuzzy logic and AI techniques. It shows improved prediction accuracy compared to conventional methods. However, limitations include reliance on specific clinical indicators and the need for extensive datasets for validation . [7]

The paper presents a neuro-fuzzy rule-based classifier, ANFIS, for detecting chronic kidney disease (CKD) using blood test results. It outperforms conventional classifiers like Multi-layer Perceptron and Support Vector Machine by achieving 3 percent to 4 percent higher accuracy. However, limitations regarding dataset size and diversity are noted [1]

The paper presents an Adaptive Neuro Fuzzy Inference System (ANFIS) model for early prediction of Chronic Kidney Disease (CKD). It achieves a prediction accuracy of 94 percent in diagnosing CKD stages. However, limitations include potential overfitting and reliance on specific datasets for training. [3]

The paper explores machine learning techniques for predicting chronic kidney disease (CKD) using patient data like blood pressure and urine values. It highlights the algorithms' high accuracy and sensitivity in early diagnosis. However, limitations regarding data diversity and algorithm generalizability are noted. [8]

The paper presents a fuzzy logic-based expert system for diagnosing chronic kidney disease (CKD), using input variables such as nephron functionality, blood sugar, and BMI to assess disease stages. The system achieved a success rate of 93.75 percent in tests, demonstrating its effectiveness in supporting doctors' diagnoses. Limitations include the inability to adapt to new data and the potential for increased complexity with more input variables, which could affect accuracy. [9]The paper presents a fuzzy logic approach to diagnose kidney disease, addressing the lack of expert availability and high consultation costs. The system successfully identified Acute Renal Failure based on symptoms. However, limitations include reliance on symptom data and potential inaccuracies without expert validation. [2] Several studies have explored machine learning techniques for CKD prediction, including decision trees, support vector machines, and neural networks. However, these models often lack the ability to explain their decisions, which is a significant limitation in healthcare. Explainable AI (XAI) techniques, such as LIME and SHAP, have been employed to interpret the decisions of complex models, offering insights into the factors contributing to the prediction. [10]

In this paper, we combine the ANFIS model, which is known for its capability to handle uncertain data, with an attention mechanism to improve the interpretability and accuracy of CKD prediction models. The fuzzy system helps capture the inherent uncertainty in medical data, while the attention mechanism enables the model to focus on the most relevant features.

3 Methodology

The proposed methodology integrates the Adaptive Neuro-Fuzzy Inference System (ANFIS) with an attention mechanism to enhance the interpretability and performance of a Chronic Kidney Disease (CKD) prediction model. This section provides a comprehensive explanation of the individual components and their integration.

3.1 Proposed Hybrid Model

The proposed model integrates the ANFIS framework with the attention mechanism and extends it with additional dense layers for final classification.Our model consists of two primary components:

- **ANFIS Layer**: The Adaptive Neuro-Fuzzy Inference System (ANFIS) uses fuzzy logic to handle the uncertainty in medical data. The model is built with a fixed number of fuzzy rules that are learned from the data. ANFIS is a hybrid model that combines the interpretability of fuzzy logic systems with the learning ability of neural networks. It uses a Sugeno-type fuzzy inference system, which models relationships in the data using fuzzy rules.

3.1.1 Fuzzy Inference System

A Sugeno-type fuzzy inference system comprises a set of fuzzy rules of the form:

$$R_i : \text{If } x_1 \text{ is } A_{i1} \text{ and } x_2 \text{ is } A_{i2} \ldots \text{ and } x_n \text{ is } A_{in}, \text{ then } y_i = w_i x + b_i \tag{1}$$

Here:

- R_i is the i-th rule in the fuzzy rule base.
- A_{ij} represents fuzzy sets defined by Gaussian membership functions.
- y_i is the output of the rule, parameterized by w_i and b_i.

1. **Fuzzification**: The input features $\mathbf{x} = [x_1, x_2, \ldots, x_n]$ are mapped to fuzzy membership values using Gaussian membership functions:

$$\mu_{A_{ij}}(x_j) = \exp\left(-\frac{(x_j - c_{ij})^2}{2\sigma_{ij}^2}\right) \tag{2}$$

   where c_{ij} is the center of the Gaussian function, and σ_{ij} is the width (spread).

2. **Rule Activation**: Each fuzzy rule R_i is activated based on the membership values:

$$\alpha_i = \prod_{j=1}^{n} \mu_{A_{ij}}(x_j) \tag{3}$$

3. **Normalization**: The activation levels are normalized to ensure they sum to 1:

$$\beta_i = \frac{\alpha_i}{\sum_{k=1}^{m} \alpha_k} \tag{4}$$

4. **Defuzzification**: The overall output is calculated as a weighted sum of the rule outputs:

$$y = \sum_{i=1}^{m} \beta_i \cdot (w_i x + b_i) \tag{5}$$

   where the weights w_i and biases b_i are learnable parameters optimized during training.

The ANFIS component in this model is designed with 12 fuzzy rules, balancing complexity and interpretability.

- **Attention Mechanism**: The attention mechanism helps the model focus on the most important features by assigning weights to each input feature based on its relevance.The attention mechanism is integrated into the ANFIS framework to dynamically focus on the most relevant fuzzy rules, thereby improving the model's interpretability and performance. [6] In traditional ANFIS, all fuzzy rules contribute equally to the final decision after normalization. However, not all rules are equally important for every input instance. The attention mechanism allows the model to learn and assign weights to the rules dynamically, emphasizing the most relevant ones.

3.1.2 Attention Weight Calculation

Given the normalized rule activations β_i from ANFIS, the attention weights a_i are computed as:

$$a_i = \frac{\exp(W \cdot \beta_i + b)}{\sum_{j=1}^{m} \exp(W \cdot \beta_j + b)} \tag{6}$$

where:

- W is a learnable weight matrix,
- b is a learnable bias vector,
- m is the number of fuzzy rules.

The softmax function ensures that the attention weights a_i sum to 1.

3.1.3 Attention Output

The attention mechanism modifies the rule activations by weighting them with the attention scores:

$$z_i = a_i \cdot \beta_i \tag{7}$$

The resulting vector $\mathbf{z} = [z_1, z_2, \ldots, z_m]$ represents the attended fuzzy rule activations.

3.1.4 Model Explainability with LIME

To ensure the interpretability of the model's predictions, the Local Interpretable Model-agnostic Explanations (LIME) framework is utilized. LIME explains individual predictions by approximating the model locally with an interpretable surrogate model, such as a linear regression or decision tree, which provides insights into the contribution of input features to the final prediction.

3.1.5 Workflow of LIME

1. **Instance Sampling:** For a given instance from the dataset, LIME perturbs the input features by generating a set of new samples around the instance, using small random variations.

2. **Model Predictions:** The proposed Hybrid ANFIS + Attention model predicts outcomes for these generated samples.

3. **Feature Importance Estimation:** A weighted interpretable model (e.g., linear regression) is fit to the perturbed data, with weights assigned based on proximity to the original instance.

4. **Explanation Generation:** The coefficients of the interpretable model are used to indicate the relative importance of input features in influencing the model's prediction.

3.1.6 Mathematical Formulation

Given an instance x, LIME creates a perturbed dataset $\mathcal{Z} = \{(x'_i, y'_i)\}$, where:

- x'_i are perturbed instances around x,

- $y'_i = f(x'_i)$, where f is the black-box model (Hybrid ANFIS + Attention).

The interpretable surrogate model g minimizes the objective:

$$\mathcal{L}(f, g, \pi_x) + \Omega(g), \tag{8}$$

where:

- $\mathcal{L}$ is the loss function comparing predictions of f and g,

- π_x is a proximity measure between x and x'_i,

- $\Omega(g)$ is a complexity measure to ensure simplicity of g.

LIME was employed post-training to evaluate the model's decision-making process for chronic kidney disease (CKD) predictions. The key steps include:

1. Selecting a test instance from the dataset.

2. Generating local explanations using the surrogate model for feature importance.

3. Visualizing the explanations to demonstrate the contribution of critical features to the prediction.

This integration of LIME highlights the transparency of the proposed method, allowing stakeholders to understand the logic behind predictions, particularly in high-stakes applications like healthcare.

The final prediction is made by combining the outputs from the ANFIS and attention layers and passing them through fully connected layers. The architecture consists of the following components:

1. **Input Layer**: Accepts normalized feature vectors $\mathbf{X} = [x_1, x_2, \ldots, x_n]$, where each $x_i \in [0, 1]$ is scaled using Min-Max normalization.

2. **Fuzzy Layer**: Implements Gaussian membership functions for fuzzification and computes the initial fuzzy rule activations α_i.

3. **Rule Activation Layer**: Computes the normalized rule activations β_i using softmax normalization.

4. **Attention Mechanism**: Processes the normalized rule activations β_i to compute attention-weighted activations z_i, focusing on the most significant fuzzy rules.

5. **Fully Connected Layers**: Two dense layers process the attention-weighted activations:

$$h_1 = \mathrm{ReLU}(W_1 \cdot \mathbf{z} + b_1), \quad h_2 = \mathrm{ReLU}(W_2 \cdot h_1 + b_2) \tag{9}$$

6. **Output Layer**: Outputs the probability of CKD using a sigmoid activation function:

$$\hat{y} = \frac{1}{1 + \exp(-(W_o \cdot h_2 + b_o))} \tag{10}$$

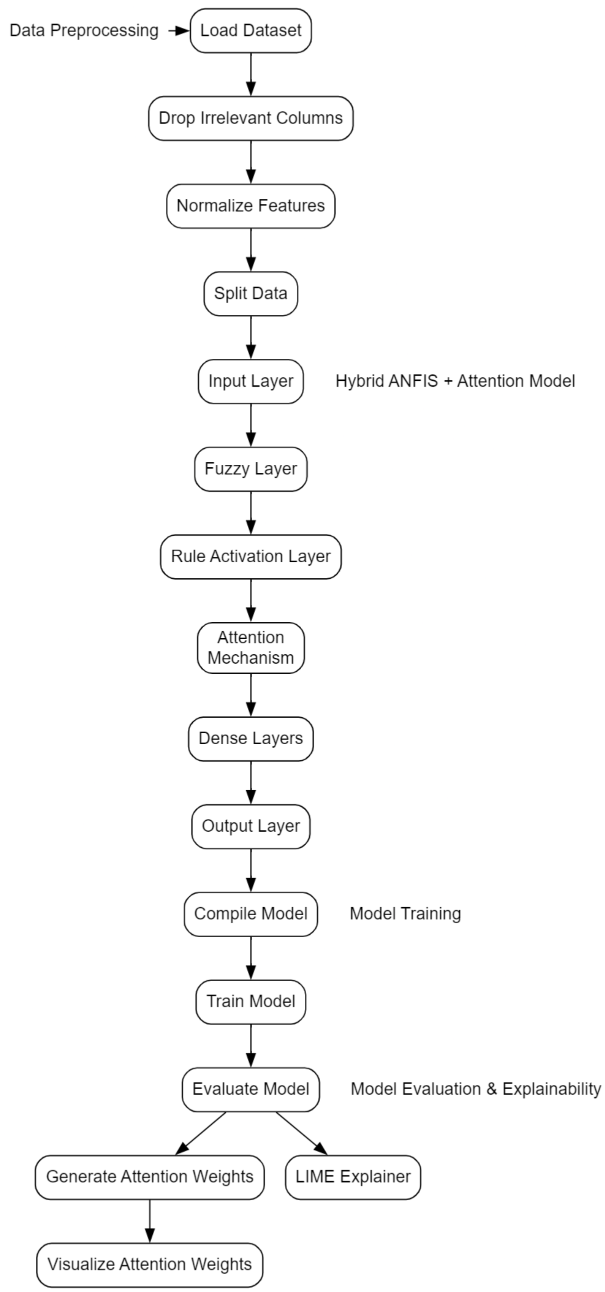

Figure 1: Flowchart of the Hybrid ANFIS + Attention Model for CKD Prediction.

3.2 Model Training

The model is trained to minimize the binary cross-entropy loss:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \tag{11}$$

where:

- y_i is the true label for the i-th instance,
- $\hat{y}_i$ is the predicted probability.

Training is optimized using the Adam optimizer with a learning rate of 1×10^{-3}. Early stopping and learning rate reduction are applied to prevent overfitting.

4 Results

In this study, we evaluated the performance of our proposed hybrid model, which integrates Adaptive Neuro-Fuzzy Inference System (ANFIS) and an attention mechanism, on the Chronic Kidney Disease (CKD) dataset. The model was trained using a set of well-established evaluation metrics: accuracy, precision, recall, and F1 score.

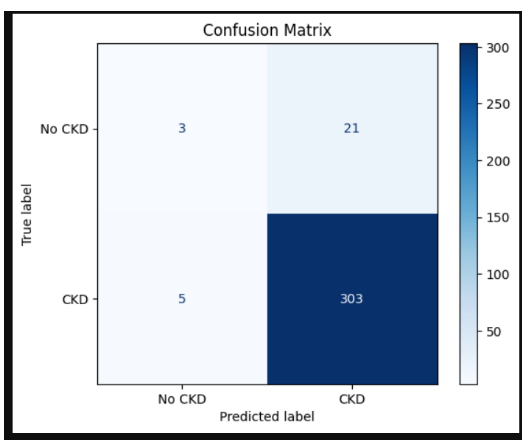

Figure 2: Confusion Matrix

The model achieved a remarkable accuracy of 95.5 percent, demonstrating its ability to correctly classify most instances in the dataset. Additionally, the precision was calculated as 0.9463, indicating that out of all the positive predictions made by the model, 94.63 percent were correct. This highlights the model's strong performance in minimizing false positives.

The recall of the model was 0.9400, meaning that the model was able to correctly identify 94 percent of all actual positive instances (CKD patients) in the dataset. This high recall value reflects the model's effectiveness in detecting chronic kidney disease, which is crucial for early diagnosis and treatment.

Finally, the F1 score, which is the harmonic mean of precision and recall, was 0.9625. This metric further confirms the balanced performance of the model, showing that it effectively maintains a trade-off between precision and recall.

Overall, the hybrid ANFIS model with attention mechanism demonstrated exceptional performance in predicting chronic kidney disease, with high accuracy and balanced precision and recall scores. These results suggest that our model is a reliable tool for supporting early detection of CKD, which can be vital for improving patient outcomes.

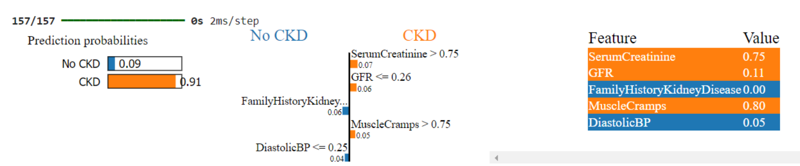

Figure 3: LIME

5 Conclusion and Future Work

5.1 Conclusion

This research proposed a novel hybrid framework combining the Adaptive Neuro-Fuzzy Inference System (ANFIS) with an attention mechanism for predicting Chronic Kidney Disease (CKD). The integration of ANFIS provided interpretability through its fuzzy logic rules, while the attention mechanism enhanced the model's ability to focus on critical features and fuzzy rules, improving prediction accuracy and efficiency. Additionally, the application of LIME (Local Interpretable Model-agnostic Explanations) further contributed to the explainability of individual predictions by highlighting the most influential features for decision-making.

The experimental results on the CKD dataset demonstrated the efficacy of the hybrid model, achieving competitive performance metrics such as high accuracy, precision, recall, and F1-score. The visualization of attention weights also illustrated the importance of specific fuzzy rules in the decision process, showcasing the model's ability to provide interpretable and actionable insights for medical diagnosis. Overall, the proposed framework successfully addressed the dual challenge of accuracy and interpretability in predictive modeling for CKD.

5.2 Future Work

While the results of this study are promising, there are several avenues for further exploration and improvement:

- **Enhanced Attention Mechanisms:** Investigate advanced attention mechanisms such as multi-head attention or self-attention to capture more complex dependencies among features and rules.

- **Broader Dataset Applications:** Extend the application of the proposed hybrid model to diverse datasets, including multi-class classification tasks in other medical domains, to evaluate its generalizability.

- **Real-Time Decision Support Systems:** Develop a practical implementation of the proposed model in a real-time decision support system for clinical use, integrating user-friendly interfaces and visualization tools.

- **Explainability Enhancement:** Expand the use of explainable AI techniques, such as SHAP (SHapley Additive exPlanations) and counterfactual explanations, to provide deeper insights into model decisions.

By pursuing these research directions, the proposed methodology can be further refined and adapted to address a wider range of challenges, contributing significantly to the field of interpretable and accurate AI-driven solutions in healthcare and beyond.

References

[1] Detection of chronic kidney disease using neuro-fuzzy rule-based classifier. 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2022.

[2] Anggun Andini, Akim Manaor Hara Pardede, and Marto Sihombing. Diagnosis of kidney disease using fuzzy logic method. *International Journal of Informatics, Economics, Management and Science*, 3(1):108–108, 2024.

[3] Komal Damodara and Anita Thakur. Adaptive neuro fuzzy inference system based prediction of chronic kidney disease. IEEE, 2021.

[4] Abdeljalil El-Ibrahimi, Sara Laghmati, Khadija Hicham, Soufiane Hamida, Bouchaib Cherradi, and Omar Bouattane. Fuzzy logic based expert system for early predicting of chronic kidney disease. 2024.

[5] Muhammad Hamid, Muhammad Saleem, and Saadia Malik. A novel approach for chronic kidney disease identification empowered with fuzzy logic. *Deleted Journal*, 4:37–52, 2024.

[6] Biomed Research International. Retracted: Fuzzy logic-based systems for the diagnosis of chronic kidney disease. *BioMed Research International*, 2024.

[7] N Pareek, Deepika Soni, and Sheshang Degadwala. Prediction of ckd using expert system fuzzy logic ai. pages 103–108, 2023.

[8] Anupam Shukla, Fabian Robin, Harshit Narvaria, and Anju Joshi. Prediction of chronic kidney disease using ml techniques. volume 5, pages 469–475, 2024.

[9] Jimmy Singla, Balwinder Kaur, Deepak Prashar, Sudan Jha, Gyanendra Prasad Joshi, Kyungyun Park, Usman Tariq, and Changho Seo. A novel fuzzy logic-based medical expert system for diagnosis of chronic kidney disease. *Mobile Information Systems*, 2020:1–13, 2020.

[10] Dr. KR Vineetha, Dr. M.S Maharajan, Mrs Bhagyashree K, and Dr. N Sivakumar. Classification of adaptive back propagation neural network along with fuzzy logic in chronic kidney disease. *e-Prime*, 2024.