Received: 30 November 2025 | Revised: 2 March 2026 | Accepted: 18 March 2026 | Published online: 7 May 2026

RESEARCH ARTICLE

Journal of Computational and Cognitive Engineering
2026, Vol. 00(00) 1–12
DOI: 10.47852/bonviewJCCE62028575

BON VIEW PUBLISHING

Explainable Fuzzy Modeling Through Antecedent Reduction and Antecedent Association in Type-1 ANFIS System

Muhammad Hamza Azam[1,2,*], Mohd Hilmi Hasan[1,2], Saima Hassan[3], Noureen Talpur[2] and Muhammad Huzaifa Azam[3]

[1] *Centre for Research in Data Science, Universiti Teknologi PETRONAS, Malaysia*

[2] *Department of Computing, Universiti Teknologi PETRONAS, Malaysia*

[3] *Institute of Computing, Kohat University of Science and Technology, Pakistan*

Abstract: The increasing demand for interpretable artificial intelligence (AI) has underscored the importance of explainable fuzzy models, particularly those based on Adaptive Neuro-Fuzzy Inference Systems. However, standard grid-based initialization often leads to exponential rule proliferation, significantly compromising model transparency and computational efficiency. To address this challenge, this study proposes an interpretable Type-1 fuzzy framework that utilizes Fuzzy C-Means clustering for data-driven membership function generation, followed by a systematic two-stage reduction strategy. The methodology integrates antecedent merging based on Euclidean distance to consolidate overlapping clusters and an activation-based pruning approach to eliminate inactive logic. Validated on the Fisher Iris dataset and Banknote Authentication dataset, the framework successfully reduced the rule base from 81 to 44 and 81 to 42 rules, achieving a 45.68% and 48.15% reduction in complexity, respectively. Crucially, this optimization enhanced classification accuracy from 76.00% to 93.33% for the Iris dataset and 44.08% to 51.13% for the Banknote Authentication dataset, demonstrating that removing redundant rules actively reduces logical noise. These results confirm that the proposed paradigm effectively balances parsimony with performance, offering a robust solution for explainable AI applications, with future extensions envisioned toward Type-2 systems for handling higher-order uncertainty.

Keywords: explainable artificial intelligence (XAI), Adaptive Neuro-Fuzzy Inference System (ANFIS), fuzzy logic system, Fuzzy C-Means clustering, antecedent reduction

1. Introduction

The increasing reliance on artificial intelligence (AI) in mission-critical applications such as healthcare, finance, and autonomous systems is compromising the interpretability and trustworthiness of AI-driven decisions. Despite the immense accuracy offered by deep learning and other black-box algorithms, their lack of explainability makes them unsuitable for applications where transparency is needed. This has led to the newly emerging field of explainable artificial intelligence (XAI) that aims at creating models that are not only accurate but human understandable as well [1]. Here, fuzzy logic-based systems have emerged as potential candidates for interpretable AI since they have the ability to express knowledge in terms of human-interpretable linguistic rules and transparent decision-making architecture [2].

Adaptive Neuro-Fuzzy Inference Systems (ANFIS) combine the neural network's learning ability with fuzzy logic humanlike reasoning. The hybrid systems are particularly valued for their balance between interpretability and adaptability. Conventional ANFIS based on Type-1 Fuzzy Logic Systems (T1FLS) has found extensive use in numerous fields for function approximation, control, and prediction [3]. But one of its major downfalls is rule explosion, a principle by which the number of fuzzy rules grows exponentially with the growth in the input dimensions or the granularity [4]. This makes the rule bases complex, and they are incapable of being explained, updated, and deployed.

T1FLS employ precise membership functions (MFs) to convert real-world inputs into linguistic terms. While this is somewhat more interpretable than black-box models, it is still marred by poor scalability and flexibility to handle high-dimensional noise or overlapping data. Second, manually optimizing or designing MFs is time-consuming and less than optimal [5]. As a result, there is an urgent need to automate MF generation and decrease the rule base without decreasing accuracy or interpretability.

In this paper, we present a comprehensive framework to enhance Type-1 ANFIS systems' scalability and explainability.

***Corresponding author:** Muhammad Hamza Azam, Centre for Research in Data Science, Universiti Teknologi PETRONAS and Department of Computing, Universiti Teknologi PETRONAS, Malaysia. Email: muhammad_17007652@utp.edu.my

Our approach uses Fuzzy C-Means (FCM) clustering for adaptive fuzzy MF generation, which is then used in fuzzy rule base initialization. To mitigate rule complexity, we use antecedent reduction based on pruning, which identifies and eliminates redundant or sparsely contributing antecedents. In addition, we conduct fuzzy rule merging to integrate semantically equivalent rules. This rigorous simplification and reduction of the antecedents guarantee the compactness and interpretability of the rule base, making it easier for the fuzzy model to be used in practice.

We innovate by jointly maximizing the interpretability and compactness of fuzzy systems in a data-driven manner. While most existing methods either focus on improving the accuracy of prediction alone or carry out rule pruning in isolation, our system preserves semantic consistency by employing fuzzy association rules. The reduced rule base not only reduces computational efforts but also delivers crisp, non-redundant linguistic rules to human analysts. This is in line directly with the goals of XAI by making model reasoning explainable and verifiable, most specifically in domains where the "why" of a prediction is as important as the prediction itself.

Despite numerous advancements in fuzzy modeling, there remains a significant gap in explainable optimization in Type-1 ANFIS systems. Present systems are often centered on heuristic rule simplification, and there isn't much focus on retaining interpretability as the model is simplified. Besides that, the lack of structure in antecedent space leads to redundant or unnecessarily complex rules that are not understandable to end-users. Our approach addresses these challenges by formalizing rule reduction and emphasizing human-centered rule abstraction in a way that the final fuzzy system is both precise and understandable.

The main contributions of this work are threefold. First, we advocate for an adaptive Type-1 MF generation using FCM to group input data adaptively and preserve semantic boundaries. Second, we perform antecedent association through merging and antecedent pruning to simplify the fuzzy rule base such that no model performance is lost. Third, we demonstrate that the resulting compact ANFIS system can still generate high prediction accuracy as compared to its original version while enhancing interpretability. The efficiency of the proposed technique is established on the Fisher Iris and Banknote Authentication datasets with reduced rule complexity and enhanced model explainability.

To better introduce the proposed framework and the contributions, the remainder of the paper is structured as follows. Section 2 provides an extended literature review of XAI, fuzzy systems, ANFIS models, and existing approaches for antecedent reduction and association. Section 3 describes the methodology in terms of using FCM for MF generation, antecedent pruning, and fuzzy rule merging. Section 4 describes the experimental setup and results by using the Fisher Iris Banknote Authentication datasets to establish the efficiency of the proposed method. Section 5 describes the implications of the findings, potential limitations, and areas for improvement. Lastly, Section 6 presents the directions of future work and ends the paper.

2. Background Study and Literature Review

The development of explainable and intelligent systems has quickly advanced over the last few years, particularly in the context of fuzzy logic and neuro-fuzzy modeling. A number of research studies have explored the potential of fuzzy logic systems (FLS) and ANFIS to solve uncertainty handling and approximate reasoning problems. However, as the data became increasingly complex in real-world usage and model explainability was increasingly necessary, researchers began to look not only at increasing accuracy but also at attempting to make the internal workings of these systems easier to understand and explain. This has led to the implementation of XAI principles along with fuzzy modeling techniques oriented around interpretability, rule simplification, and semantic transparency. Detailed analysis of influential ideas, methodological innovations, and recent breakthroughs in this field is necessary to understand the state of the art and identify areas for enhancement. The subtopics below present a comprehensive overview of the most critical constituents and methodologies for use with FLS and explainable neuro-fuzzy modeling.

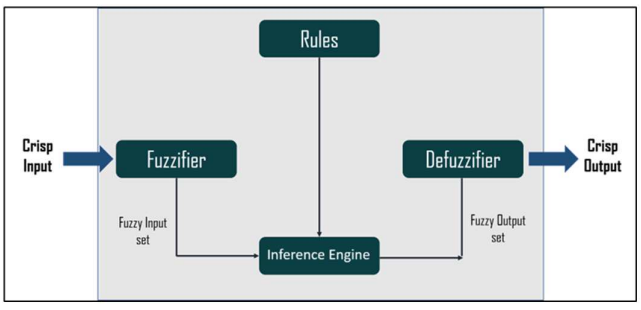
2.1. Fundamentals of fuzzy logic systems

FLS, first proposed by Lotfi A. Zadeh in 1965, transformed how uncertain and imprecise information was managed in computational systems. Unlike classical binary logic, where variables have to be exactly true or false (0 or 1), fuzzy logic permits variables to have values within a range between 0 and 1. The characteristic allows real-world phenomena that are inherently fuzzy or imprecise, like "hot weather," "moderate speed," or "low risk," to be expressed. A fuzzy definition like this allows the transition between values to be done gradually and increasingly humanlike responses, and thus, FLS saw tremendous success in imprecise and uncertain environments.

A conventional FLS consists of three key components: fuzzification, inference engine, and defuzzification. The crisp numerical values are, during fuzzification, transformed into membership degrees in existing fuzzy sets by means of MFs [6]. The inference engine operates on the basis of a set of IF–THEN rules defining the system's behavior by the use of fuzzy logic operators. Rules are generally formulated by linguistic variables (e.g., "IF temperature is high THEN fan speed is fast"). The defuzzification subsequently generates one crisp output from the fuzzy output set by implementing the centroid method primarily or by implementing other defuzzification techniques. The architecture of the fuzzy system is illustrated in Figure 1.

One of Type-1 FLS's traditional advantages is that of built-in interpretability. Since fuzzy rules are expressed using linguistic variables, there is close similarity to how human decisions are expressed, and systems are transparent as well as auditable [7]. The stakeholders can look at the fuzzy rules and MFs themselves to see decisions being made. For instance, for a medical decision-making application, a rule like "IF bp is high AND cholesterol is high THEN risk is severe" is interpretable as well as clinically

Figure 1
Architecture of fuzzy logic system



interpretable [8]. The feature makes Type-1 fuzzy systems candidates to be deployed to XAI applications, especially safety-critical ones like healthcare, finance, and autonomous systems [9].

Nevertheless, in spite of their inherent practicality, FLS are not issue-free. As input variable and fuzzy term values increase, the rule base grows exponentially, a curse called the "curse of dimensionality" or rule explosion. This may mean hundreds of scores or thousands of fuzzy rules that complicate the system to work with and understand. This calls for the use of methods like antecedent reduction, rule association, and fuzzy clustering that reduce model complexity without impacting its accuracy as well as interpretability [10]. They are a prerequisite of rendering fuzzy systems interpretable and scalable to applications in the real world.

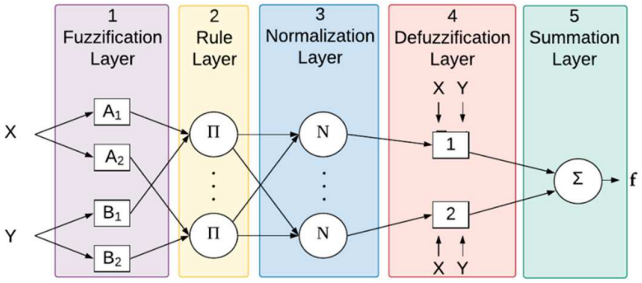
2.2. Adaptive Neuro-Fuzzy Inference Systems (ANFIS)

ANFIS, introduced by Jang in 1993, is a hybrid intelligent system that includes training of neural networks along with clear fuzzy logic rule representation [11]. ANFIS has a multilayer structure that is commonly composed of five layers that execute specific operations like fuzzification, rule evaluation, normalization, and defuzzification. The training process utilized by ANFIS is a hybrid process that includes the least squares estimation and gradient descent to enable automatic adjustments of MFs as well as rule parameters with training data [12]. The adaptability that is obtained that way makes ANFIS strong enough to approximate complex nonlinear functions with very good accuracy while still being able to characterize a transparent rule-based structure.

An important strength of ANFIS is that it can balance data-friendliness with explainer-friendliness. Every rule in ANFIS corresponds to one particular fuzzy region in the input space and can be explained as language (e.g., "IF temperature is high AND humidity is low THEN fan speed is high"). The simplicity of this fuzzy rule base makes ANFIS highly accessible to application development in XAI, where understanding decision logic in a model is key [13]. The use of fuzzy logic is such that one can embed expertise in a domain directly in one's rule base, and all one must do to fit this to data is learn one's parameter set (even learn this set using Bayesian-type updating). There are dozens of application examples that have been able to adequately demonstrate ANFIS's capabilities in application domains such as this one (medical diagnosis, control systems, finance prediction, environmental modeling) as a function of this trade-off between accuracy and interpretable comprehensibility [14].

Although it is strong, when it comes to high-dimensional problems, it is restricted. The more input variables or the more linguistic labels, the greater the number of fuzzy rules, and an exponential increase of the latter is an issue typically referred to as "the curse of dimensionality" [15]. The resulting rule explosion creates redundancy or contradiction among rules, causing increased computational load and untangling issues of model explanation. Additionally, due to the overlapping nature of fuzzy MFs, sometimes there are fuzzy rule firings that are not clear, and as a result, model output explanation becomes even more challenging [16]. Due to the latter, there is recent work that has investigated the application of antecedent pruning, rule simplification, and MF generation optimization (e.g., using clustering schemes like FCM) to make ANFIS models transparent and efficient enough for demanding application purposes. The conceptual design of ANFIS is presented in Figure 2.

Figure 2
ANFIS architecture diagram



2.3. Explainable artificial intelligence (XAI) in fuzzy modeling

XAI intends to make decision-making within artificial systems transparent and interpretable to human end-users. With growing incorporation of AI systems into mission-critical application areas such as healthcare, finance, and autonomous transport systems, there are also growing requirements for models whose decisions are interpretable and can be trustworthy [17]. The common traditional black-box type of models such as deep neural networks do not usually give meaningful insights concerning decision-making, and this gives rise to concern with regard to responsibility and fairness issues [18]. The FLS, however, have inherent interpretable reasoning with human-interpretable linguistic variables by rule systems. The fuzzy systems are thus considered to be a natural candidate for schemes of XAI.

Fuzzy systems employ IF–THEN rules built out of linguistic terms such as "Low," "High," or "Moderate," just as humans reason. The rules are directly linked to comprehensible ideas such that end-users can monitor, validate, and even debug a line of reasoning behind a system's output. On data-fused versions like ANFIS, this clarity is still maintained while being data-dynamically capable of learning. For example, after training, an ANFIS model may output this rule: "IF temperature is High and humidity is Low THEN fan_speed = 0.8," which is numerically correct as well as linguistically interpretable. The end effect is that those who know a domain are able to examine a line of reasoning free of special-purpose tools/post hoc interpretation layers [19]. Interpretability suffers with excessively complex fuzzy systems. High numbers of fuzzy rules, overlapping MFs, or inconsistent rule bases blur as opposed to enhance interpretation. Experiments have been capable of greatly shedding light upon systems by compacting fuzzy rules and compressing antecedents that are unimportant or redundant [20]. Beyond this, MF semantic clarity and linguistic term quality also contribute to how interpretable a fuzzy model is. On behalf of ends that are XAI, one must ensure that one ends up with a compact, unambiguously clear, semantically informative rule base, thus highlighting antecedent reduction and association as important methods to fuzzy modeling.

2.4. Managing complexity in fuzzy rule bases: challenges, reduction, and association techniques

When FLS are scaled up to real-world, high-dimensional data, overly complex rule bases are obtained as solutions. With ANFIS, this kind of complexity is mainly induced by combinatorial rule explosion as a result of additional input variables and

MFs [21]. With five variables as inputs with three MFs per variable, up to 243 candidate fuzzy rules are produced by the model. Other than imposing undue computational burdens, such rule sets also deter model interpretation to an extent that stakeholders are not readily able to make sense of decisions being made [22]. Such overlapping and compact rule sets are not manageable to employ within application sectors like healthcare, finance, and safety-critical systems that need to be explained and transparent.

In response to such challenges, antecedent reduction schemes have been introduced by previous authors. Rule pruning is one such typical technique that practically removes rules contributing the minimum to model performance or the minimum to activate the input space in a systematic way. The technique retains the most informative decision paths with minimum noise within the inference process. Antecedent merging is also one technique that identifies and merges rules with corresponding analogous input conditions as generalized rules. Apart from retaining rules to a minimum by retaining the total count to a minimum, this technique also removes redundant linguistic patterns and thereby maximizes transparency. Feature selection by mutual information or by correlation analysis can also retain the considered input variables to a minimum within the fuzzy inference system (FIS) and thereby retain the overall rule base to a minimum. The techniques are employed to construct lightweight, transparent fuzzy models easily verifiable and auditable within real-world applications [23]. Rule association and rule merge methods reduce model complexity further by extracting dependencies as well as rules from rule bases using data mining ideas such as extracting frequent patterns [24]. Such methods extract common antecedent patterns and pair them with linguistic aggregation. The Apriori algorithm and fuzzy measures of correlation, for instance, are able to extract sets of antecedents that are responsible for all occurrences together and are thus able to replace such sets with compact, generalized rules that are less descriptive. Such association-related methods reduce semantic overlap to a minimum and maximize the accuracy of a model that is compatible with interpretable modeling targets. The trade-off between accuracy and compactness forms the basis of compact fuzzy systems designed for interpretable AI systems [25].

Against this background, this research paper adopts a hybrid approach that employs a rule pruning and antecedent merge technique to minimize fuzzy rule base compactness. The broad approach is to keep the overall system predictive capacity undisturbed while achieving immensely enhanced interpretable capacity with a less expansive semantically interpretable rule base as well. The approach, by focusing on such antecedent reduction and association-based methods, complements an explainer-friendly as well as human-interpretable fuzzy modeling approach that is rather indispensable to allow transparent decision-making under high-stakes conditions.

3. Methodology

The proposed method is an explainable fuzzy modeling framework based on Type-1 ANFIS with antecedent reduction and association techniques for minimizing fuzzy rule bases. It has five major steps in the process: data preprocessing, fuzzy partitioning through FCM, ANFIS modeling, rules pruning and antecedent merging, and finally, assessment through interpretability and accuracy metrics. The approach is targeted at improving the interpretability of fuzzy models without sacrificing predictive performance, particularly in medium-scale data such as the Fisher Iris dataset and Banknote Authentication dataset.

3.1. Data preparation and preprocessing

To validate the proposed fuzzy rule reduction framework, the classic Fisher Iris dataset and Banknote Authentication dataset were utilized as the benchmark. The details of the datasets are as follows:

3.1.1. Fisher Iris dataset

The dataset consists of 150 samples distributed equally across three species of Iris flowers: *Iris setosa*, *Iris versicolor*, and *Iris virginica*. Each sample contains four numerical features:

1) Sepal Length (×1)
2) Sepal Width (×2)
3) Petal Length (×3)
4) Petal Width (×4)

3.1.2. Banknote Authentication dataset

The Banknote Authentication dataset contains 1375 samples, each representing an image of a banknote-like specimen. The goal of the dataset is to distinguish between genuine and forged banknotes based on visual texture information extracted from these images.

Each sample is described using four numerical features that are computed from wavelet-transformed images. These features capture different aspects of the image texture: the *variance (VWTI)* reflects how much the pixel intensities vary, *skewness (SWTI)* shows whether the intensity distribution is balanced or tilted, *kurtosis (KWTI)* indicates how sharp or flat the distribution is, and *entropy (EI)* measures the overall randomness or complexity of the image.

The images were captured using an industrial inspection camera, similar to those used in professional printing environments, to ensure consistent and high-quality image acquisition. Both real and counterfeit banknote samples were included in the data collection process. Overall, the dataset provides a compact yet informative representation with four input attributes and a class label, making it well-suited for experimenting with classification models, fuzzy systems, and XAI techniques.

3.2. Fuzzy C-Means clustering for membership function generation

The second phase is automated construction of fuzzy MFs from FCM clustering. As opposed to manual specification of MFs, the FCM algorithm gives data-driven and objective partitioning of the input space. FCM assigns points in every input feature to pre-specified clusters from which triangular MFs are obtained.

This MF implementation through clustering guarantees that the fuzzy sets align with the natural distribution of data; hence, the model accuracy and interpretability both become better. The proposed algorithm used to approximate Type-1 Fuzzy Triangular MF using FCM is presented in Algorithm 1:

Algorithm 1: Triangular Membership Function Approximation

1. Choose (e, m, iter, ε, a, b, c)
2. Find initial cluster centers
3. ITERATE

For t = 1 to iter

CALCULATE

$$u_{ik,\,t} = \left[\sum_{j=1}^{e} \left(\frac{|x_k - v_{i,\,t-1}|_A}{|x_k - v_{j,\,t-1}|_A} \right)^{1/(m-1)} \right]^{-1}$$

CALCULATE $v_{i,\,t} = \frac{\sum_{k=1}^{N} (u_{ik,\,t})^m \, X_k}{\sum_{k=1}^{N} (u_{ik,\,t})^m}$

If error $= |v_t - v_{t-1}| \leq \varepsilon$

Next t

4. *U-Matrix* and cluster center "*e*" are calculated.

5. Calculate parametric values for triangular MF.

$$\begin{aligned} a &= \alpha - (\beta \times \gamma) \\ b &= \alpha \\ c &= \alpha + (\beta \times \gamma) \end{aligned}$$

6. Triangular MF set

$$\mu_F(x; a, b, c) = \max\left(\min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0\right)$$

7. Use parametric values $a < b < c$ to generate triangular MF

Here, a and c denote the lower and upper bounds of the triangular MF, while b represents its peak value. The parameters α and γ are computed from the U-matrix and the corresponding cluster centers, whereas β is treated as a constant. The value of β is determined empirically based on experimental analysis and simulation results.

3.3. ANFIS training and initial rule base construction

The Type-1 ANFIS architecture integrates fuzzy logic inference with the learning capabilities of neural networks to optimize MF parameters and consequent weights. Utilizing the data-driven fuzzy sets derived via FCM clustering, the system initializes the rule base by computing the full Cartesian product of the input antecedents. This exhaustive process generates a rule for every logical combination of the MFs, ensuring comprehensive coverage of the feature space. However, this combinatorial approach inevitably leads to the "curse of dimensionality," resulting in an exponentially large rule base that necessitates subsequent reduction to maintain computational efficiency.

3.4. Antecedent association and rule merging

Following the initial rule generation, the system comprised a comprehensive but redundant set of rules, many of which occupied adjacent or spatially overlapping regions in the feature space. To address this, a geometric association-based merging method was employed to consolidate these entities into generalized patterns. Unlike complex semantic measures, this approach utilized Euclidean distance as the primary metric to quantify the spatial proximity between the four-dimensional antecedent centers of the generated rules.

The merging process was executed in two logical steps. First, a pairwise distance matrix was computed for all rule centers, and a similarity threshold of 0.2 was applied to identify "connected components," or groups of rules indistinguishable within that spatial tolerance. Second, Rule Generalization was performed, where each identified cluster was consolidated into a single representative rule. This strategy minimizes structural redundancy and enhances the model's linguistic interpretability by avoiding fine-grained distinctions that do not add explanatory value.

3.5. Antecedent reduction through rule pruning

Even after the initial generation and merging phases, the rule base may still contain "null" or weak rules that occupy regions of the feature space devoid of empirical data. To mitigate this and enhance model parsimony, a validity-based pruning algorithm is utilized to assess the actual contribution of each rule to the inference process. The core of this method involves Activation Analysis, where the cumulative activation strength for every rule is computed by summing its firing strength across the entire training dataset. This metric effectively quantifies the degree to which a specific rule participates in explaining the input data, distinguishing between dominant logical structures and inactive noise.

Based on this analysis, a Thresholding and Elimination strategy is applied. A predefined validity threshold is established to identify statistically insignificant rules; any rule with a total activation strength falling below this value is deemed inactive or redundant. These rules are systematically removed from the rule base, ensuring that only those with significant empirical support are retained. This elimination process serves to drastically reduce computational overhead and logical complexity without compromising important decision patterns, ultimately yielding a transparent and human-understandable rule base.

3.6. Explainability evaluation

To rigorously assess the transparency of the proposed framework, a dual-faceted evaluation strategy comprising both quantitative and qualitative metrics was employed. Quantitatively, the reduction in rule count served as the primary indicator of interpretability; minimizing the rule base from an initial exhaustive set to a compact subset directly correlates with reduced cognitive load for human analysts. Qualitatively, linguistic simplicity and semantic distinctness were validated through the antecedent merging phase, which ensured that the retained rules occupied spatially distinct regions of the feature space (Euclidean distance >= 0.2). This geometric separation guarantees that the final rule base utilizes consistent and non-overlapping linguistic labels, thereby preventing the confusion typically caused by inconsistent or highly similar antecedent definitions.

3.7. Performance metrics

To quantitatively validate the predictive capability of the system, classification accuracy was employed as the definitive performance metric. Defined as the ratio of correctly classified samples to the total number of instances in the dataset, this metric provided a direct measure of the model's decision-making precision. It served as the benchmark for comparative analysis, allowing for a rigorous assessment of the trade-off between model complexity (rule count) and generalization ability. By monitoring accuracy shifts between the initial exhaustive model and the final reduced model, the framework ensured that the enhancement in interpretability did not compromise the system's ability to correctly identify the Iris species or fake banknotes.

4. Experiments and Results Discussion

The explainable fuzzy modeling framework proposed was empirically tested on the Fisher Iris dataset and Banknote

Authentication dataset. The primitive Type-1 FIS model was initialized with four input variables, and each was partitioned into three fuzzy sets using the FCM algorithm to obtain a total of 81 fuzzy rules. These rules included all of the input membership combinations and hence made the model highly expressive but also complex and difficult to understand. Through some simplification, a lot of inapplicable rules were eliminated by rule reduction based on antecedent merging and activation strength pruning. As a result, the rule base was successfully reduced from 81 to 44 and 81 to 42 rules of substance that were effective, demonstrating a significant reduction in complexity while retaining interpretability. Each of these steps is explained below.

4.1. Fuzzy membership function generation using FCM

Each of the four input features was clustered using FCM into three clusters, generating semantically meaningful, overlapping fuzzy sets. These centers were used to construct triangular MFs, which were symmetric and interpretable.

The generated fuzzy MFs are depicted in Figure 3, which shows the fuzzy MFs generated for each of the four input features of the Iris and Banknote Authentication datasets. Each curve represents a triangular fuzzy set derived from the FCM clustering results.

Each input variable was normalized to the numeric range [0, 1] prior to clustering. The FCM algorithm partitioned each input into three clusters ($c = 3$), and the cluster centroids were used to parameterize triangular MFs $\mu(x; a, b, c)$ satisfying $a < b < c$. These functions represent Low, Medium, and High linguistic terms. These MFs form the fuzzification layer in the inference system and provide an interpretable linguistic structure for rule generation.

4.2. Initial rule base construction

The initial rule base was established by partitioning the four input variables using the FCM algorithm to derive triangular MFs. By generating all logical combinations of these fuzzy sets (34), an exhaustive set of 81 rules was constructed, ensuring comprehensive coverage of the feature space prior to the application of reduction techniques.

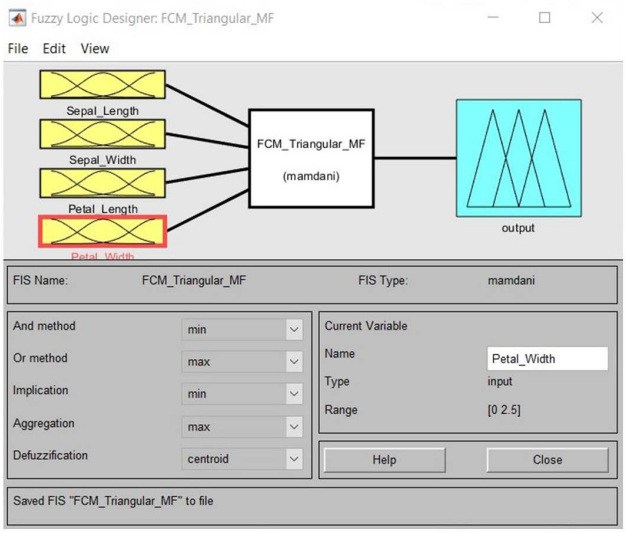
4.3. Antecedent merging based on similarity

The rule merging targets the reduction of model redundancy by identifying fuzzy rules that govern highly similar regions of the input space. This process is predicated on the geometric interpretation of rule antecedents, where each rule is treated as a centroid in the multidimensional feature space, in this case, a four-dimensional space corresponding to the Iris and Banknote Authentication dataset attributes. The similarity between any pair of rules is quantified using the Euclidean distance between these centroids. A predefined similarity threshold of 0.2 serves as the decision boundary; pairs of rules with an inter-center distance less than this value are deemed semantically equivalent and candidates for consolidation.

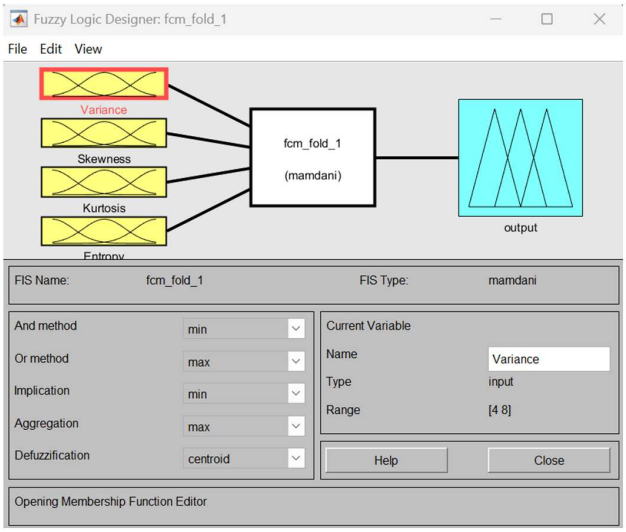
Upon satisfying the merging criterion, the algorithm consolidates the identified pair into a single, generalized rule, reducing the rule set from 81 to 68 and 81 to 72 for the Iris and Banknote Authentication datasets, respectively. This is achieved by averaging the antecedent parameters of the parent rules and retaining the consequent class that possesses the higher accumulated firing strength, ensuring the dominant logical inference is preserved. This topological reduction strategy effectively lowers the computational burden without degrading the system's ability to discriminate between classes. The spatial relationships between the rule centers and the connectivity indicating merged clusters are graphically illustrated in Figure 4.

Figure 3
Triangular MF FIS through FCM for (a) Iris dataset and (b) Banknote authentication dataset

(a)



(b)



4.4. Rule pruning based on activation strength

To eliminate low-contributing rules, each rule's average activation strength across the training data was calculated. Rules with a mean activation below a defined threshold (0.03) were considered insignificant and pruned.

Figure 5 illustrates the rule activation strengths for all 68 rules for the Iris dataset and 72 rules for the Banknote Authentication dataset, with a red dashed line showing the pruning threshold.

Figure 4
Principal component analysis (PCA) of rule merging for antecedent reduction for (a) Iris dataset and (b) Banknote authentication dataset

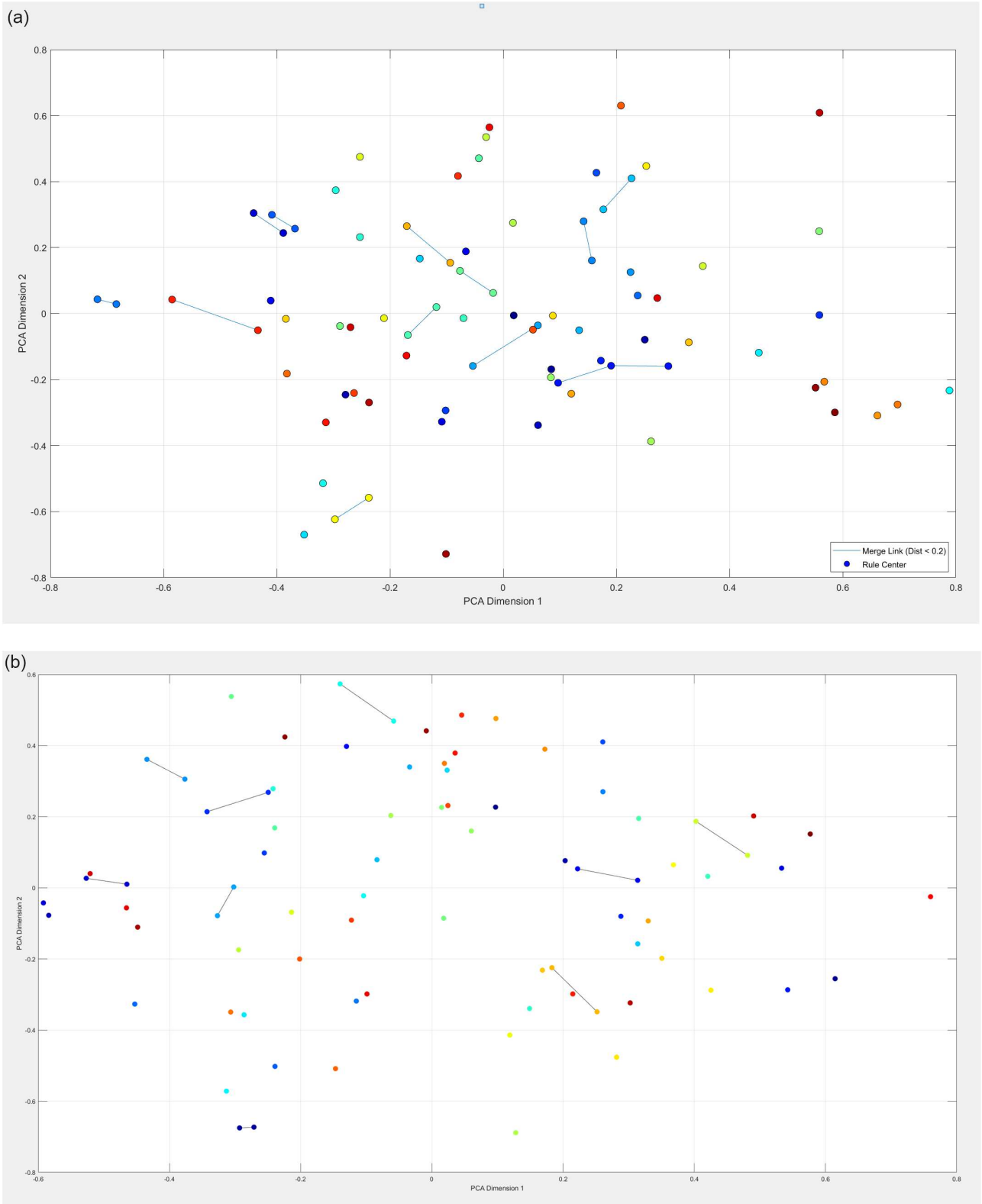


Figure 5

Distribution of rule activation strengths highlighting pruned rules for (a) Iris dataset and (b) Banknote authentication dataset

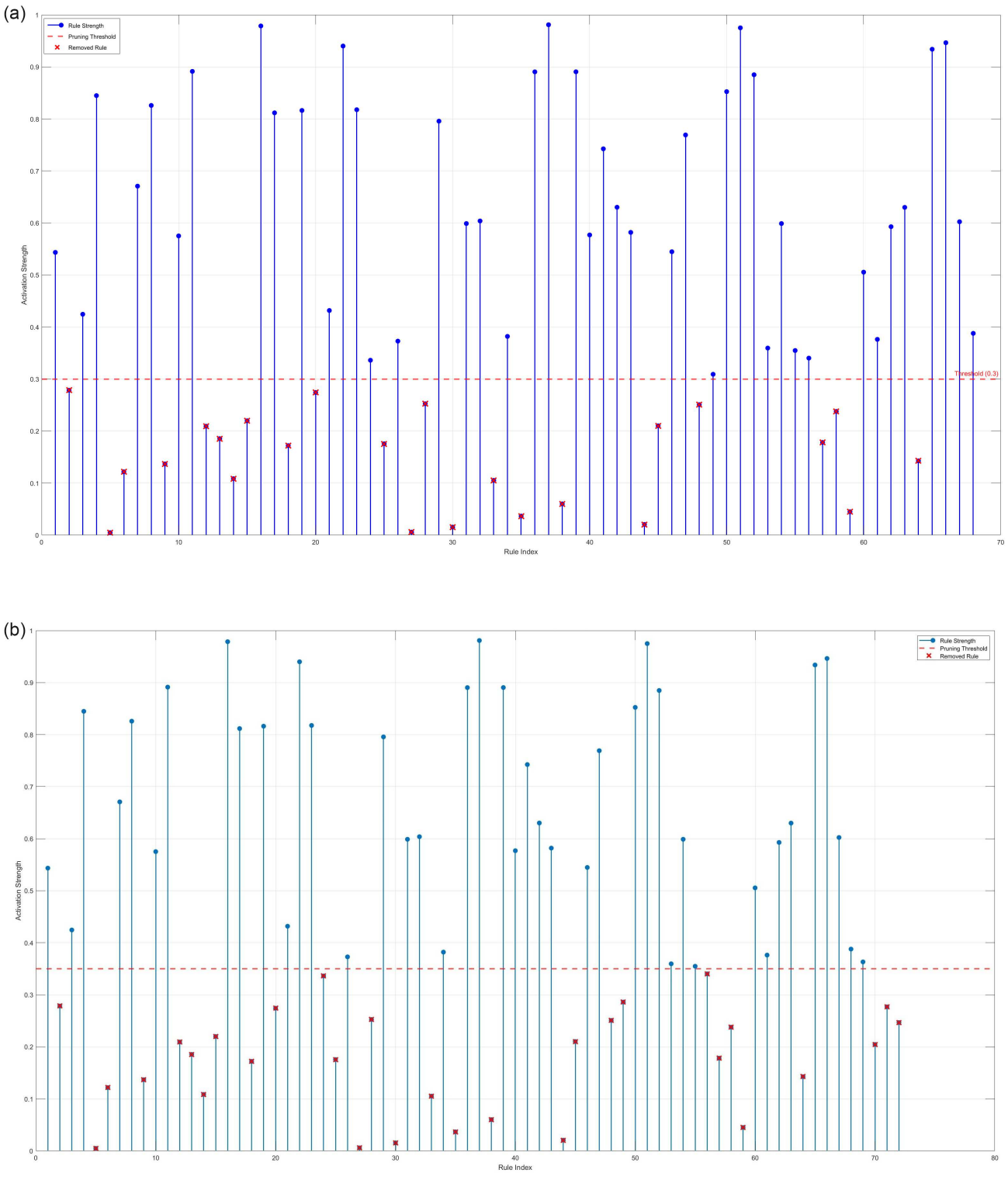
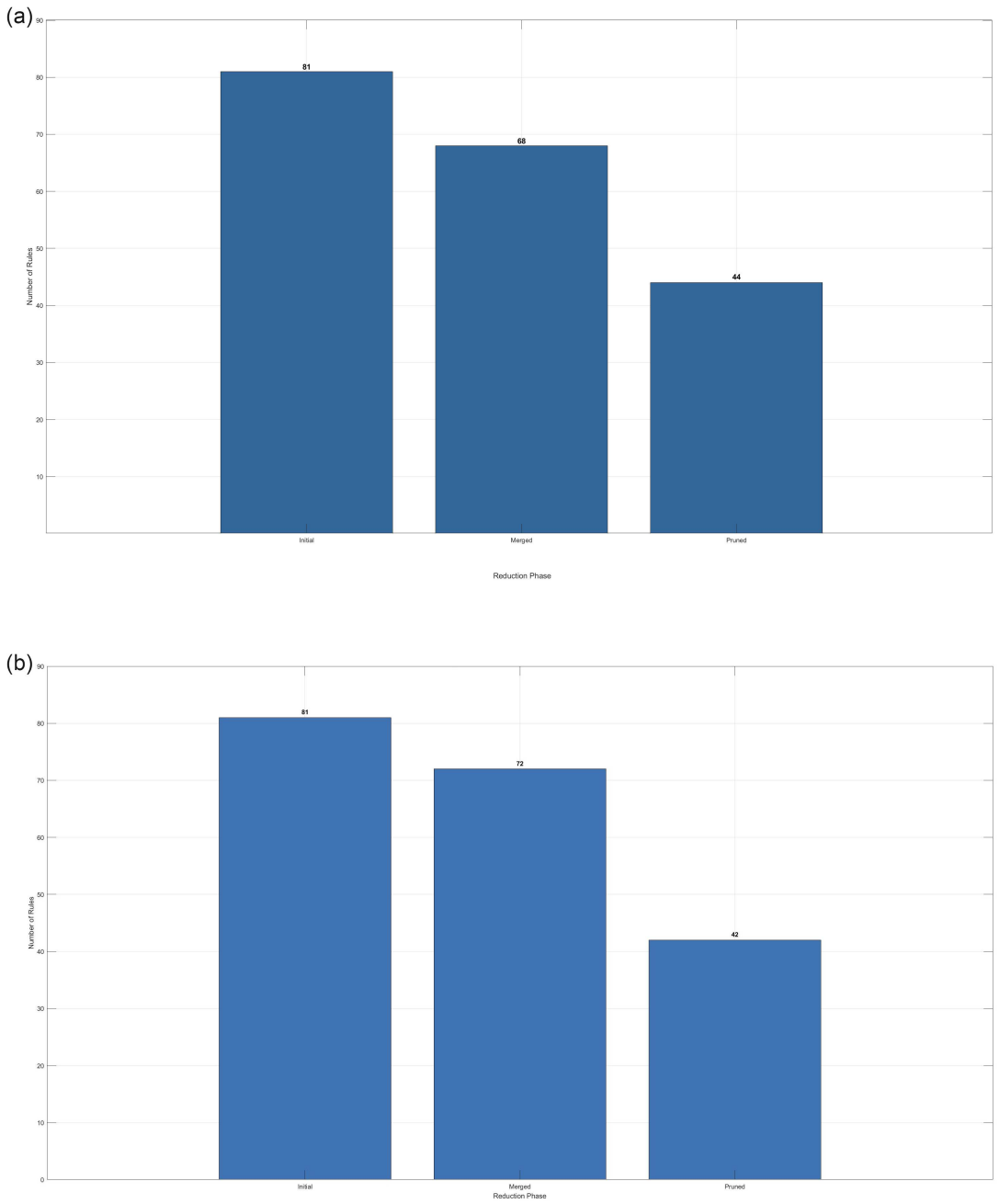


Figure 6
Effect of antecedent merging and Pruning on rule base size for (a) Iris dataset and (b) Banknote authentication dataset



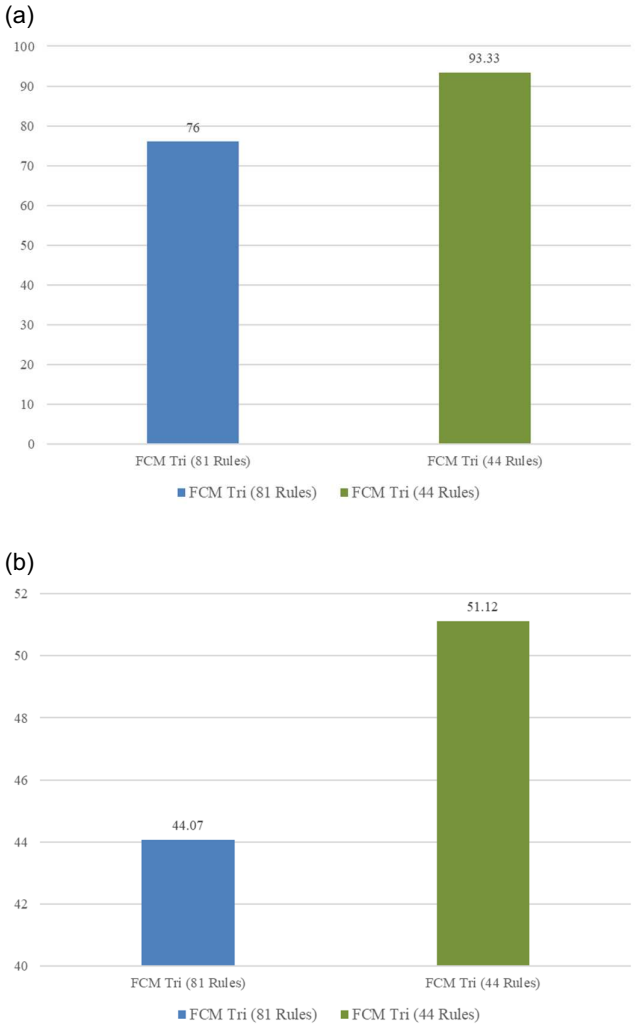
Rules falling below this line were removed, reducing the rule base from 68 to 44 and 72 to 42, respectively.

This step significantly reduces the complexity of the model without sacrificing the semantic structure.

Figure 6 illustrates the sequential reduction of model complexity across the three experimental phases. Starting with an initial exhaustive set of 81 rules, the system first underwent antecedent merging, which consolidated spatially similar clusters to reduce the count to 68 and 72 rules for the Iris and Banknote Authentication datasets, respectively. Subsequently, the validity-based pruning phase eliminated inactive or weak rules, resulting in a final compact rule base of 44 and 42 rules, respectively. This stepwise optimization represents a nearly 45.68% and 48.15% overall reduction in system complexity, respectively, while retaining the most significant logical structures.

The impact of rule reduction on system performance is presented in Figure 7. The initial model, burdened by 81 redundant rules, achieved a baseline accuracy of 76.00% for the Iris dataset, while 44.08% baseline accuracy for the Banknote Authentication dataset. In contrast, the optimized model with 44 rules demonstrated a significant improvement, reaching 93.33% accuracy for the Iris dataset, and the optimized model with 42 rules demonstrated a significant improvement, reaching 51.12% accuracy for the Banknote Authentication dataset. This confirms that eliminating noise and contradictory logic not only reduced model complexity but also enhanced its generalization capability.

Figure 7
Comparison of dataset performance before and after rule reduction for (a) Iris dataset and (b) Banknote authentication dataset



These reductions significantly enhance interpretability, reduce computational overhead, and simplify model explanation, all of which align with XAI principles. In the initial phase, antecedent association analysis based on Euclidean distance was applied to identify spatial redundancy, resulting in an immediate reduction from 81 to 68 and 81 to 72 rules for Iris and Banknote Authentication datasets, respectively. Subsequently, the pruning process based on cumulative activation strength eliminated inactive logic, further refining the rule base to 44 and 42 significant rules for the Iris and Banknote Authentication datasets, respectively. Crucially, this optimization was not merely syntactic; it directly enhanced predictive performance, boosting classification accuracy from 76.00% to 93.33% for the Iris dataset and 44.07% to 51.12% for the Banknote Authentication dataset by filtering out contradictory noise. The successful application of the merging mechanism demonstrates its importance for scalability, as future high-dimensional datasets are expected to produce even denser overlapping antecedents, where this fuzzy union-based generalization will be essential.

5. Conclusion

The proposed research introduced an explainable fuzzy modeling paradigm centered on antecedent merging and rule base pruning in Type-1 FIS. Utilizing FCM clustering, an initial comprehensive rule base of 81 fuzzy rules was generated. To address inherent redundancy, a systematic two-stage reduction framework was implemented: antecedent merging based on Euclidean distance first consolidated overlapping clusters to 68 and 72 rules for the Iris and Banknote Authentication datasets, respectively, followed by a pruning technique based on cumulative activation that refined the system to 44 and 42 significant rules, respectively. This combined approach achieved an overall 45.68% and 48.15% reduction in model complexity for the Iris and Banknote Authentication datasets, respectively.

Contrary to the assumption that model compression compromises performance, empirical testing on the Fisher Iris and Banknote Authentication dataset revealed a substantial increase in classification accuracy, rising from 76.00% in the initial model to 93.33% in the reduced model for the Iris dataset and 44.07% in the initial model to 51.12% in the reduced model for the Banknote Authentication dataset. This performance inversion demonstrates that the removed rules were contributing primarily to logical noise and inference ambiguity. Consequently, the proposed merging and pruning procedures served not only as compression methods but also as robust feature selection tools that clarified the decision boundaries.

While the framework demonstrates strong interpretability, it inherently assumes crisp membership grades typical of Type-1 models, which limits its applicability in domains with highly non-stationary or noisy input distributions. Future extensions toward Interval Type-2 FIS are essential to model secondary uncertainty. Integrating these higher-order fuzzy sets would broaden the framework's robustness, allowing it to handle more complex uncertainties while maintaining the benefits of the proposed reduction strategies.

Ultimately, this outcome emphasizes the advantage of multi-stage rule reduction for XAI, particularly for real-world applications where transparency and traceability are paramount. The reduced rule base not only makes decision logic more transparent to human observers but also offers significant opportunities for integrating efficient, interpretable fuzzy models into low-resource or real-time control environments.

6. Future Work

A primary avenue for future research is the extension of this methodology to Interval Type-2 Fuzzy Inference Systems (IT2-FIS). Unlike the current Type-1 approach, IT2 systems model secondary uncertainty via footprints of uncertainty, making them highly effective for noisy or non-stationary environments. Since Type-2 systems typically generate significantly larger and more complex rule bases, they are ideal candidates for the proposed advanced pruning and merging techniques. Future work will focus on adapting the antecedent merging logic to handle secondary MFs, ensuring that enhanced robustness does not come at the cost of model interpretability.

To validate scalability and resistance to the "curse of dimensionality," the framework will be evaluated on high-dimensional, real-world datasets in domains such as energy optimization, medical diagnosis, and financial forecasting. Supporting this expansion, research will also explore hybrid reduction architectures that combine numerical pruning with semantic similarity

measures and ontology-based rule association. By integrating statistical validity with linguistic analysis, the framework aims to ensure that rules remain not only accurate but also semantically distinct when applied to complex, multi-variable problems.

Finally, to bridge the gap between algorithmic optimization and end-user adoption, efforts will focus on developing interactive explainability dashboards. Integrating visual representations of the reduction process, such as the PCA projections and activation plots demonstrated in this study, will facilitate broader industrial adoption. These tools will enable stakeholders to visualize rule consolidation and validate decision boundaries in real time, thereby fostering greater trust, transparency, and traceability in AI-derived decision-making.

Funding Support

This research is supported by the Ministry of Higher Education Malaysia under the Fundamental Research Grant Scheme (FRGS), reference no. FRGS/1/2022/ICT02/UTP/02/1 (015MA0-160) and TotalEnergies EP Malaysia (015MD0-165).

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in the UCI Machine Learning Repository at https://archive.ics.uci.edu/dataset/53/iris and https://archive.ics.uci.edu/dataset/267/banknote+authentication.

Author Contribution Statement

Muhammad Hamza Azam: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Mohd Hilmi Hasan:** Conceptualization, Resources, Supervision, Project administration, Funding acquisition. **Saima Hassan:** Writing – review & editing, Supervision. **Noureen Talpur:** Validation, Formal analysis, Writing – review & editing. **Muhammad Huzaifa Azam:** Writing – review & editing, Visualization.

References

[1] Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., ..., & Hussain, A. (2024). Interpreting black-box models: A review on explainable artificial intelligence. *Cognitive Computation*, *16*(1), 45–74. https://doi.org/10.1007/s12559-023-10179-8

[2] Huang, Y., & Wagner, C. (2025). On the potential of fuzzy integral-based decision-level fusion when the fuzzy measure is informed by densities alone. In *2025 IEEE International Conference on Fuzzy Systems*, 1–6. https://doi.org/10.1109/FUZZ62266.2025.11152126

[3] Jafarzade, N., Kisi, O., Yousefi, M., Baziar, M., Oskoei, V., Marufi, N., & Mohammadi, A. A. (2023). Viability of two adaptive fuzzy systems based on fuzzy c means and subtractive clustering methods for modeling Cadmium in groundwater resources. *Heliyon*, *9*(8), e18415. https://doi.org/10.1016/j.heliyon.2023.e18415

[4] Gökmen, Ö. B., Güven, Y., & Kumbasar, T. (2025). FAME: Introducing fuzzy additive models for explainable AI. In *2025 IEEE International Conference on Fuzzy Systems*, 1–6. https://doi.org/10.1109/FUZZ62266.2025.11152211

[5] Ali, M. A., Mekhilef, S., Yusoff, N., & Abd Razak, B. (2022). Laser simulator logic: A novel inference system for highly overlapping of linguistic variable in membership functions. *Journal of King Saud University - Computer and Information Sciences*, *34*(10), 8019–8040. https://doi.org/10.1016/j.jksuci.2022.07.017

[6] Lima, J. F., Patiño-León, A., Orellana, M., & Zambrano-Martinez, J. L. (2025). Evaluating the impact of membership functions and defuzzification methods in a fuzzy system: Case of air quality levels. *Applied Sciences*, *15*(4), 1934. https://doi.org/10.3390/app15041934

[7] Azam, M. H., Hasan, M. H., Malik, A. A., Hassan, S., & Abdulkadir, S. J. (2022). Energy price forecasting through novel fuzzy type-1 membership functions. *Computers, Materials & Continua*, *73*(1), 1799–1815. https://doi.org/10.32604/cmc.2022.028292

[8] Murad, N. Y., Hasan, M. H., Azam, M. H., Yousuf, N., & Yalli, J. S. (2024). Unraveling the black box: A review of explainable deep learning healthcare techniques. *IEEE Access*, *12*, 66556–66568. https://doi.org/10.1109/ACCESS.2024.3398203

[9] Azam, M. H., Hasan, M. H., Murad, N. Y., & Patah, E. A. B. (2024). Transparency in AI: A review of explainable artificial intelligence techniques. In *2024 8th International Conference on Computing, Communication, Control and Automation*, 1–5. https://doi.org/10.1109/ICCUBEA61740.2024.10774981

[10] Kanth, M. V. (2025). Using fuzzy logic for improving model interpretability in machine learning classifiers. *Global Journal of Engineering Innovations & Interdisciplinary Research*, *5*(1), 1–5. https://doi.org/10.33425/3066-1226.1069

[11] Jang, J.-S. R. (1993). ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics*, *23*(3), 665–685. https://doi.org/10.1109/21.256541

[12] Jang, J.-S. R., & Sun, C.-T. (1995). Neuro-fuzzy modeling and control. *Proceedings of the IEEE*, *83*(3), 378–406. https://doi.org/10.1109/5.364486

[13] Alonso Moral, J. M., Castiello, C., Magdalena, L., & Mencar, C. (2021). *Explainable fuzzy systems: Paving the way from interpretable fuzzy systems to explainable AI systems*. Switzerland: Springer. https://dx.doi.org/10.1007/978-3-030-71098-9

[14] Samnioti, A., & Gaganis, V. (2023). Applications of machine learning in subsurface reservoir simulation—A review—Part I. *Energies*, *16*, 6079. https://doi.org/10.20944/preprints202307.0630.v1

[15] Wang, J., Wu, Y., Huang, X., Zhang, C., & Nie, F. (2024). Projected fuzzy c-means clustering algorithm with instance penalty. *Expert Systems with Applications*, *255*, 124563. https://doi.org/10.1016/j.eswa.2024.124563

[16] Gu, X., Han, J., Shen, Q., & Angelov, P. P. (2023). Autonomous learning for fuzzy systems: A review. *Artificial*

Intelligence Review, *56*(8), 7549–7595. https://doi.org/10.1007/s10462-022-10355-6

[17] Gunning, D., Vorm, E., Wang, Y., & Turek, M. (2021). *DARPA's explainable AI (XAI) program: A retrospective*. Authorea Preprints.

[18] Kumar, S., Sarraf, S., Kar, A. K., & Ilavarasan, P. V. (2023). A study of eXplainable artificial intelligence: A systematic literature review of the applications. In P. K. Singh, S. T. Wierzchoń, W. Pawłowski, A. K. Kar, & Y. Kumar (Eds.), *IoT, big data and AI for improving quality of everyday life: Present and future challenges: IOT, data science and artificial intelligence technologies* (pp. 243–259). Springer. https://doi.org/10.1007/978-3-031-35783-1_14

[19] Mendel, J. M., & Bonissone, P. P. (2021). Critical thinking about explainable AI (XAI) for rule-based fuzzy systems. *IEEE Transactions on Fuzzy Systems*, *29*(12), 3579–3593. https://doi.org/10.1109/TFUZZ.2021.3079503

[20] Ferchichi, A., Abbes, A. B., Barra, V., & Farah, I. R. (2025). Trustworthy AI for spatio-temporal forecasting via counterfactual causality. In *2025 IEEE 37th International Conference on Tools with Artificial Intelligence*, 10–17. https://doi.org/10.1109/ICTAI66417.2025.00010

[21] Cao, J., Zhou, T., Zhi, S., Lam, S., Ren, G., Zhang, Y., . . . , & Cai, J. (2024). Fuzzy inference system with interpretable fuzzy rules: Advancing explainable artificial intelligence for disease diagnosis—A comprehensive review. *Information Sciences*, *662*, 120212. https://doi.org/10.1016/j.ins.2024.120212

[22] Yogeesh, N., Mohammad, S. I., Raja, N., Chetana, R., William, P., Vasudevan, A., . . . , & Alshurideh, M. T. (2025). From crisp to fuzzy: A comparative review of statistical and fuzzy approaches to problem solving. *Applied Mathematics & Information Sciences*, *19*(3), 647–658. http://dx.doi.org/10.18576/amis/190313

[23] Zhang, K., Shao, T., Sun, Y., Xu, X., Zhang, X., Zhou, X., . . . , & Huang, S. (2026). Interpretable research of fuzzy methods: A literature survey. *Information Fusion*, *126*, 103524. https://doi.org/10.1016/j.inffus.2025.103524

[24] Tian, M.-W., Alattas, K., El-Sousy, F., Alanazi, A., Mohammadzadeh, A., Tavoosi, J., . . . , & Skruch, P. (2022). A new short term electrical load forecasting by type-2 fuzzy neural networks. *Energies*, *15*(9), 3034. https://doi.org/10.3390/en15093034

[25] Hegazi, M. O., Almaslukh, B., & Siddig, K. (2023). A fuzzy model for reasoning and predicting student's academic performance. *Applied Sciences*, *13*(8), 5140. https://doi.org/10.3390/app13085140

How to Cite: Azam, M. H., Hasan, M. H., Hassan, S., Talpur, N., & Azam, M. H. (2026). Explainable Fuzzy Modeling Through Antecedent Reduction and Antecedent Association in Type-1 ANFIS System. *Journal of Computational and Cognitive Engineering*. https://doi.org/10.47852/bonviewJCCE62028575