

Article

GRS-ANFIS: A Gate-Network-Based Role-Separated ANFIS for Interpretable Classification

Jeong Heon Lee ^{1,†}, Sangwook Kim ^{2,3,†} and Sungmoon Jeong ^{2,3,*} 

¹ Department of Artificial Intelligence, Kyungpook National University, Daegu 41566, Republic of Korea; ljh1112@knu.ac.kr

² AI-Driven Medical Innovation Center, Kyungpook National University Hospital, Daegu 41944, Republic of Korea; kim@knu.ac.kr

³ Department of Medical Informatics, School of Medicine, Kyungpook National University, Daegu 41944, Republic of Korea

* Correspondence: jeongsm00@gmail.com

† These authors contributed equally to this work.

Abstract

Explainable artificial intelligence is increasingly needed in high-stakes tabular classification, where predictions should be accurate, auditable, and easy to inspect. We propose GRS-ANFIS, a role-separated neuro-fuzzy model that decomposes inference into a Primary module for main decision formation and a Complementary module for targeted correction. During differentiable training, sigmoid gate values are applied only to consequent coefficients, while the antecedent part receives the original input without soft masking. After each stage, the learned gates are binarized into hard routing masks that define discrete antecedent and consequent subsets for module-specific fine-tuning. The Complementary module is restricted to variables not selected by the Primary module, yielding explicit role separation and disjoint variable usage across modules. To support stable learning in high-dimensional settings, all ANFIS-family models use the same HTSK-style firing computation. Experiments on four tabular benchmarks show that GRS-ANFIS achieves competitive predictive performance while maintaining compact, role-separated rule structures; rule-count compactness is clear, whereas the unified Nauck/HFSi interpretability values are dataset- and variant-dependent. Boundary-focused analysis further shows that the Complementary module mainly improves difficult, low-confidence samples through targeted correction.

Keywords: explainable artificial intelligence; ANFIS; neuro-fuzzy systems; role-separated architecture; interpretable classification

MSC: 68T07



Academic Editor: Francisco Rodrigues Lima-Junior

Received: 4 April 2026

Revised: 13 May 2026

Accepted: 15 May 2026

Published: 18 May 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Explainable artificial intelligence (XAI) has become increasingly important in high-stakes decision-making, where predictions should be accurate, understandable, and easy to inspect [1,2]. Many such problems are naturally represented as tabular data, including credit scoring, risk assessment, and clinical decision support. In these settings, users often ask not only what a model predicts, but also why the prediction was made, preferably in a form that can be reviewed and checked.

Although modern machine learning models often show strong predictive performance on tabular data, many of them remain difficult to interpret. Post-hoc explanation methods

such as LIME and SHAP can highlight influential features for individual predictions [3,4]. However, feature attribution alone does not directly provide clear decision logic in an explicit rule form. This limitation becomes more important when users need explanations that are easy to verify and communicate.

Neuro-fuzzy systems are a useful intrinsic alternative because they express decisions through fuzzy IF-THEN rules while still allowing parameter learning from data. In particular, the adaptive neuro-fuzzy inference system (ANFIS) combines fuzzy rule bases with trainable membership functions and consequent parameters [5]. Because of this rule-based structure, ANFIS is often viewed as a promising model for interpretable tabular learning.

However, a standard TSK-ANFIS has important limitations in modern tabular settings. As the number of input features increases, the rule base can become too large and difficult to manage. This rule explosion problem reduces both computational efficiency and interpretability. In addition, several rules may be activated at the same time, which can make it difficult to connect a final prediction to a small and representative set of rules. Prior work has also shown that TSK fuzzy systems may have difficulty learning from high-dimensional inputs unless suitable firing or defuzzification strategies are used [6]. In our study, we adopt HTSK-style firing computation as a shared learning condition for all ANFIS-family models, so that the main comparison focuses on architectural differences rather than on trainability alone.

To address these scalability problems, prior studies proposed hierarchical or modular ANFIS architectures that decompose a high-dimensional input space into multiple low-dimensional fuzzy subsystems [7,8]. Such designs can reduce rule explosion and improve parameter efficiency by limiting the dimensionality handled by each module. However, in many cases, these intermediate modules are introduced mainly for tractability, and their outputs do not always correspond to semantically clear or human-meaningful roles. As a result, hierarchical decomposition alone does not necessarily provide an explanation in which one part of the model forms the main decision and another part refines it only when needed.

Several other studies have also used feature selection, sparsity constraints, or evolutionary optimization to reduce model complexity [6,9–11]. These approaches can improve compactness and sometimes predictive performance, but they still do not explicitly organize inference into clearly different functional roles. In many real-world decisions, a few key factors establish the main decision context, while other factors provide additional refinement only in uncertain or exceptional cases.

Motivated by this perspective, we propose GRS-ANFIS (Gate-network-based Role-Separated Adaptive Neuro-Fuzzy Inference System), a neuro-fuzzy framework that explicitly separates inference into two modules. The Primary module learns the main decision from a sparse feature subset, while the Complementary module learns a targeted correction from the remaining variables with the Primary module fixed. Instead of softly masking raw inputs, the model applies learnable global gate values only in the consequent layer during differentiable training. After thresholding, the resulting binary masks define hard antecedent and consequent subsets for each module. This design provides a structured explanation of the final prediction as a primary decision plus a complementary correction.

To make this separation stable, GRS-ANFIS is trained in two stages. The Primary module is optimized first with consequent-only soft gating and then converted into a hard-routed subset model. The Complementary module is optimized next on the variables not selected by the Primary module and is refined under the same soft-to-hard routing procedure. This design encourages each module to focus on a different role and supports a compact rule structure.

This paper does not claim that GRS-ANFIS alone resolves the trainability challenges of high-dimensional ANFIS. Rather, under a shared HTSK-based learning condition, we investigate whether role-separated fuzzy inference offers a better balance among predictive performance, structural compactness, and interpretability. We evaluate GRS-ANFIS on four public tabular classification benchmarks spanning low-to-high dimensional settings. For ANFIS, H-ANFIS, GRS-ANFIS, SVM, and the tree-based and additive tabular baselines (Random Forest, XGBoost, LightGBM, CatBoost, and EBM), we report the full preprocessed input representation for each dataset, whereas GA-ANFIS and PSO-ANFIS additionally include feature-selection effects as part of their optimization pipelines. In addition to Accuracy and F1 score, we report rule-based interpretability indices and conduct a boundary-focused analysis to assess whether the Complementary module mainly benefits difficult, low-confidence samples.

The main contributions of this paper are as follows:

1. We propose GRS-ANFIS, a role-separated neuro-fuzzy architecture that divides inference into a Primary module for main decision formation and a Complementary module for targeted correction, with explicit feature-role assignment through hard routing masks.
2. We introduce a two-stage training strategy that uses consequent-only soft gating during differentiable optimization and hard antecedent/consequent routing during refinement, while enforcing complementary feature usage after the Primary module is fixed.
3. Under a shared HTSK-based learning condition, we show that GRS-ANFIS achieves competitive predictive performance while maintaining a compact and role-separated rule structure across multiple tabular benchmarks, with interpretability assessed mainly through rule count and the unified Nauck/HFSi index.
4. We provide a boundary-focused analysis of low-confidence samples to examine the correction role of the Complementary module.

2. Related Work

2.1. Neuro-Fuzzy Systems and ANFIS for Tabular Learning

Fuzzy inference systems provide decision logic in the form of IF–THEN rules, which can be easier to communicate and audit than purely numeric representations. Neuro-fuzzy models combine fuzzy rules with gradient-based learning, enabling rule parameters and membership functions to be adapted from data. ANFIS is a representative framework that implements this idea using a Sugeno-type fuzzy model with trainable antecedent membership functions and consequent parameters [5].

Recent fuzzy-model interpretability studies further emphasize that fuzzy structure should not be treated as an automatic guarantee of interpretability. Fuzzy models can provide semantic transparency through rules, membership functions, and graph-like causal structures, but interpretability may degrade when rule bases expand, when fuzzy components are tightly coupled with complex learners, or when data-driven links encode spurious associations [12,13]. This supports our conservative framing: GRS-ANFIS is evaluated as a role-separated neuro-fuzzy model with explicit rules and sparsity evidence, rather than as a universally interpretable model by virtue of being fuzzy.

Despite this appeal, conventional (non-hierarchical) ANFIS faces well-known challenges in high-dimensional tabular settings. As the number of input variables grows, rule construction based on grid-like combinations of membership functions can suffer from the curse of dimensionality, causing the number of required rules and parameters to increase rapidly and yielding excessively large and unstructured rule bases (rule explosion) [6]. Moreover, overlapping membership functions can lead to simultaneous activation

of multiple rules. As a result, similar predictions may be produced by different combinations of rules, which blurs the mapping from a prediction to a concise and representative rule-based explanation.

2.2. Hierarchical and Modular ANFIS Variants

To address scalability and complexity, many studies propose hierarchical fuzzy systems and modular neuro-fuzzy architectures [7,8,14]. A common strategy is to decompose a high-dimensional input space into smaller subsets processed by multiple low-dimensional modules, and then combine intermediate outputs through higher-level aggregation. Such designs can reduce computational cost, alleviate rule explosion, and improve parameter efficiency by limiting the dimensionality handled by each fuzzy module.

However, in many hierarchical designs the hierarchy is introduced primarily for tractability or expressiveness, and intermediate subsystem outputs may not correspond to physically or semantically meaningful variables [8,14]. As a result, it can be difficult to interpret each intermediate subsystem, and explanations may still rely on aggregating contributions across modules rather than attributing a prediction to modules with explicit semantic roles. Therefore, while hierarchical structures can improve tractability, they do not necessarily define a stable explanatory unit that clearly separates a primary driver from a conditional modifier.

2.3. Feature Selection and Sparsity in Fuzzy Modeling

Another line of research addresses the curse of dimensionality in neuro-fuzzy modeling by reducing effective input dimensionality and suppressing redundant rules through feature selection, rule pruning, and sparsity-inducing constraints. Evolutionary and swarm-based optimization methods have been widely used as wrappers around ANFIS to search for compact configurations. For example, genetic algorithm (GA) based schemes can select informative feature subsets and then train an ANFIS model on the reduced input space, improving parsimony and interpretability by restricting the variables involved in the learned rule base [9]. Particle swarm optimization (PSO) has similarly been integrated with ANFIS to optimize model parameters or rule construction settings, often improving predictive performance while reducing the need for manual tuning in high-dimensional problems [10]. In addition to search-based methods, regularization techniques (e.g., ℓ_1 -type penalties) have been applied to encourage sparse parameterizations and compact rule bases by discouraging dense feature usage and unnecessary rules.

Beyond selecting features externally, some approaches embed differentiable gating directly into the learning procedure to perform integrated feature selection and rule extraction. A representative strategy is to apply learnable gate parameters in the consequent part of TSK/ANFIS models, where each input variable's contribution in the linear consequent is scaled by a continuous gate during training and then thresholded for pruning after optimization [11]. Consequent-only gating can be attractive in high-dimensional settings because it provides a stable and differentiable mechanism for suppressing irrelevant variables without altering antecedent membership functions or rule firing patterns, which can otherwise make optimization more brittle. However, because the antecedent structure remains unchanged, such gating primarily controls how strongly variables influence the output once rules fire, rather than simplifying which conditions activate rules or assigning explicit semantic roles to intermediate components.

Our method follows this stable consequent-only gating principle during soft optimization, but adds a subsequent hard-routing refinement stage so that the selected variables explicitly determine both antecedent participation and consequent evaluation in the role-separated modules.

Although global gates are related to mixture-of-experts routing [15], neural attention/gating mechanisms, and differentiable selection frameworks such as DSelect-k [16], GRS-ANFIS uses a different explanatory unit. The gates are sample-independent and module-specific, binarized into disjoint hard masks, and are embedded into a fuzzy IF–THEN rule system. Thus, the contribution is not gating alone, but the conversion of learned gates into a stable Primary/Complementary feature-role partition that determines both antecedent rule participation and consequent evaluation.

Overall, feature selection and sparsity methods can substantially reduce model size and improve readability, but compactness alone does not guarantee a structured explanation. In particular, when multiple components compete over overlapping features or when selection is expressed only as coefficient suppression, the resulting explanation may remain ambiguous about which feature group establishes the Primary decision versus which group provides conditional refinement. This limitation motivates mechanisms that explicitly assign distinct feature roles and separate the inference process into interpretable functional components.

3. Materials and Methods

Throughout this section, we use $\mathbf{x} \in \mathbb{R}^D$ to denote the input vector, where x_d is the d -th input variable, D is the number of input variables, R is the number of fuzzy rules, and r indexes fuzzy rules. We use $w_r(\mathbf{x})$ and $\bar{w}_r(\mathbf{x})$ for the un-normalized and normalized firing strengths, respectively, while $f_r(\mathbf{x})$ denotes the linear consequent function of rule r . For input dimension d , the j -th fuzzy set is denoted by $A_{d,j}$, and $\mu_{A_{d,j}}(x_d)$ denotes the corresponding membership degree. For module-specific formulas, we use $A_{r,d}^P$ and $A_{r,d}^C$ to denote the antecedent fuzzy set used by rule r on input dimension d in the Primary and Complementary modules, respectively.

3.1. Overall Conceptual Comparison of Rule Organization

Figure 1 summarizes how rule organization and explanation structure differ across the three ANFIS-based models considered in this paper.

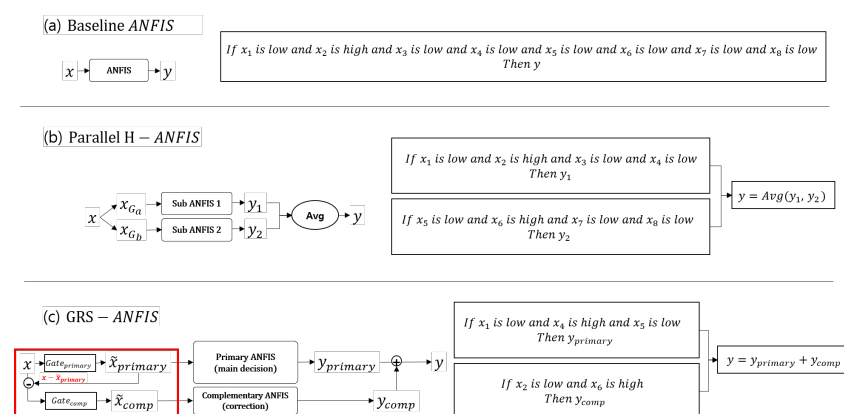


Figure 1. Conceptual comparison of rule organization and explanation structure in (a) Baseline ANFIS, (b) Parallel H-ANFIS, and (c) GRS-ANFIS. Baseline ANFIS uses full-input rules over the entire feature space; Parallel H-ANFIS shortens rules by feature-group decomposition but does not enforce explicit module roles. GRS-ANFIS separates inference into Primary and Complementary modules with explicit role assignment, yielding a summed module score $\tilde{y} = y_{\text{primary}} + y_{\text{comp}}$ and a more structured rule-level explanation. The red box in panel (c) highlights the gate-based role-routing mechanism: the Primary gate forms $\tilde{\mathbf{x}}_{\text{primary}} = \mathbf{g}_{\text{primary}}^{\text{hard}} \odot \mathbf{x}$, and the Complementary gate operates only on the remaining feature space, $(\mathbf{1} - \mathbf{g}_{\text{primary}}^{\text{hard}}) \odot \mathbf{x}$, to form $\tilde{\mathbf{x}}_{\text{comp}}$. Thus, the Complementary module is restricted to variables not selected by the Primary module.

3.2. Overview and Motivation

GRS-ANFIS combines three key design ideas to address the limitations of traditional TSK-ANFIS in high-dimensional settings:

1. **Parallel Primary–Complementary Decomposition:** Instead of a standard TSK-ANFIS, we employ two separate ANFIS modules that partition the learning task into role-separated pathways: the Primary module learns global patterns, while the Complementary module learns fine-grained corrections.
2. **Learnable Global Gates with Consequent-Only Soft Gating:** Each module uses a trainable global gate-logit vector. During soft training, the corresponding sigmoid gate values modulate only consequent coefficients, avoiding direct distortion of raw inputs and antecedent memberships.
3. **Stage-wise Soft-to-Hard Routing:** We employ a two-stage training procedure in which each module is first learned with consequent-only soft gating and then refined under fixed hard routing masks. The Primary mask is frozen before Complementary training, and the Complementary module is restricted to variables not selected by the Primary module.

3.3. Global Gate Parameterization and Hard Role Gating

To mitigate the curse of dimensionality and enforce role separation, GRS-ANFIS uses two learnable gate-logit vectors, one for the Primary module and one for the Complementary module. Unlike input-masking approaches, the soft gates are not multiplied with the raw input before fuzzification. During soft training, the antecedent part always receives the original input $\mathbf{x}$, whereas the gate values are applied only in the consequent layer.

3.3.1. Primary and Complementary Gate Parameters

We parameterize feature selection using two trainable gate-logit vectors, $\mathbf{a}_{\text{primary}} \in \mathbb{R}^D$ and $\mathbf{a}_{\text{comp}} \in \mathbb{R}^D$, shared across all samples. Soft gate values are obtained by the element-wise logistic sigmoid:

$$\mathbf{g}_{\text{primary}} = \text{sigmoid}(\mathbf{a}_{\text{primary}}), \quad \mathbf{g}_{\text{comp}} = \text{sigmoid}(\mathbf{a}_{\text{comp}}). \quad (1)$$

For the Primary module, the soft-gated consequent of rule r is

$$f_{r,\text{primary}}^{\text{soft}}(\mathbf{x}) = p_{r,0}^{\text{primary}} + \sum_{d=1}^D g_{\text{primary},d} p_{r,d}^{\text{primary}} x_d. \quad (2)$$

After soft training, the learned Primary gate is binarized as

$$g_{\text{primary},d}^{\text{hard}} = \mathbb{I}[g_{\text{primary},d} \geq \tau_{\text{primary}}]. \quad (3)$$

3.3.2. Complementary Gating Under the Primary–Complementary Constraint

In Stage 2, only variables not selected by the Primary hard gate are available to the Complementary module. Accordingly, the soft Complementary gate is defined as

$$\tilde{\mathbf{g}}_{\text{comp}} = (\mathbf{1} - \mathbf{g}_{\text{primary}}^{\text{hard}}) \odot \mathbf{g}_{\text{comp}}. \quad (4)$$

The soft-gated Complementary consequent of rule r is

$$f_{r,\text{comp}}^{\text{soft}}(\mathbf{x}) = p_{r,0}^{\text{comp}} + \sum_{d=1}^D \tilde{g}_{\text{comp},d} p_{r,d}^{\text{comp}} x_d. \quad (5)$$

After Stage 2, we threshold the Complementary gate to obtain the final hard mask:

$$g_{\text{comp},d}^{\text{hard}} = \mathbb{I}[g_{\text{comp},d} \geq \tau_{\text{comp}}], \quad \mathbf{g}_{\text{comp}}^{\text{hard}} = (\mathbf{1} - \mathbf{g}_{\text{primary}}^{\text{hard}}) \odot \mathbf{g}_{\text{comp}}^{\text{hard}}. \quad (6)$$

In this study, we use $\tau_{\text{primary}} = \tau_{\text{comp}} = 0.5$, the natural binarization threshold of the sigmoid gate, which is equivalent to thresholding the corresponding gate logits at zero. During gradient-based optimization, soft gate values are used only in the consequent layer. After each stage, hard masks are applied to define discrete antecedent and consequent subsets for routing, fine-tuning, and interpretability reporting.

3.4. Stage-Wise Training

GRS-ANFIS is trained in two stages to enforce specialization and sparse feature roles.

3.4.1. Stage 1: Primary Module Training

Let $\mathbf{a}_{\text{primary}} \in \mathbb{R}^D$ denote the learnable Primary gate logits, and let

$$\mathbf{g}_{\text{primary}} = \sigma(\mathbf{a}_{\text{primary}}) \quad (7)$$

be the corresponding soft Primary gate.

In the soft phase of Stage 1, the original input $\mathbf{x}$ is fed to the antecedent part of the Primary ANFIS, while the soft gate $\mathbf{g}_{\text{primary}}$ is applied only to the consequent part. The resulting soft Primary output is denoted by

$$\mathbf{y}_{\text{primary}}^{\text{soft}} = \text{ANFIS}_{\text{primary}}^{\text{soft}}(\mathbf{x}; \mathbf{g}_{\text{primary}}), \quad (8)$$

where the superscript “soft” indicates that consequent-only soft gating is used.

The soft Primary training objective is

$$L_{S1}^{\text{soft}} = \text{Loss}(\mathbf{y}_{\text{primary}}^{\text{soft}}, \mathbf{y}) + \lambda_{\text{primary}} \|\mathbf{g}_{\text{primary}}\|_1. \quad (9)$$

Here, the ℓ_1 penalty encourages sparsity in the learned soft Primary gate during differentiable training.

After soft training, we binarize the learned Primary gate using the threshold τ_{primary} to obtain the fixed hard mask

$$\mathbf{g}_{\text{primary}}^{\text{hard}} = \mathbb{I}[\mathbf{g}_{\text{primary}} \geq \tau_{\text{primary}}]. \quad (10)$$

The resulting hard-routed Primary output is denoted by

$$\mathbf{y}_{\text{primary}}^{\text{hard}} = \text{ANFIS}_{\text{primary}}^{\text{hard}}(\mathbf{x}; \mathbf{g}_{\text{primary}}^{\text{hard}}), \quad (11)$$

where the superscript “hard” indicates that the fixed binary Primary mask is applied to both the antecedent and consequent parts.

The hard Primary refinement objective is

$$L_{S1}^{\text{hard}} = \text{Loss}(\mathbf{y}_{\text{primary}}^{\text{hard}}, \mathbf{y}). \quad (12)$$

Because the hard mask is fixed after thresholding, no gate regularization term is used in this refinement stage. Therefore, Stage 1 hard refinement fine-tunes the Primary ANFIS under the fixed routing pattern induced by $\mathbf{g}_{\text{primary}}^{\text{hard}}$.

Figure 2 illustrates the complete Stage 1 procedure. The left panel shows consequent-only soft training of the Primary module, and the right panel shows thresholding of the learned gate followed by hard-routing refinement under the fixed Primary hard mask.

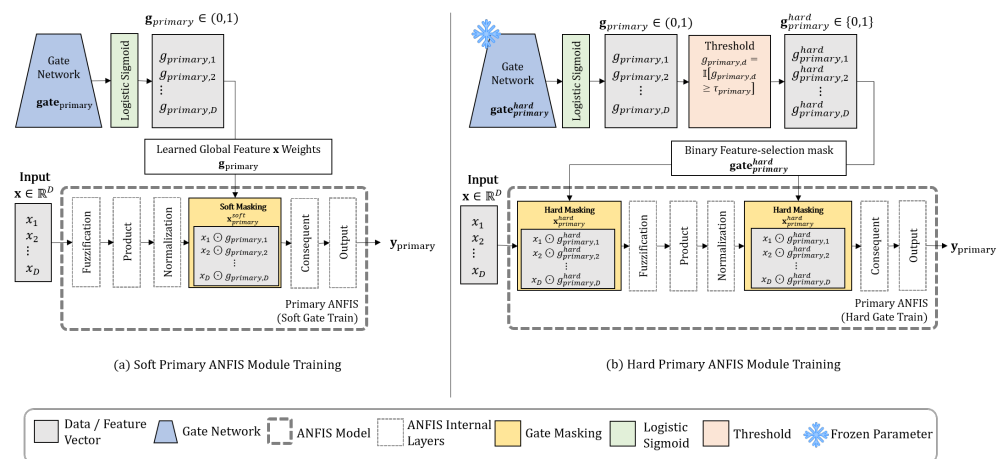


Figure 2. Stage 1 training workflow for the Primary module. The left panel shows consequent-only soft gating, and the right panel shows hard-routing refinement after thresholding the learned Primary gate.

3.4.2. Stage 2: Complementary Module Training

Let $\mathbf{a}_{\text{comp}} \in \mathbb{R}^D$ denote the learnable Complementary gate logits, and let

$$\mathbf{g}_{\text{comp}} = \sigma(\mathbf{a}_{\text{comp}}) \quad (13)$$

be the corresponding soft Complementary gate.

In Stage 2, the Complementary antecedent receives the hard-routed input subset

$$\mathbf{x}_{\text{comp}} = (\mathbf{1} - \mathbf{g}_{\text{primary}}^{\text{hard}}) \odot \mathbf{x}. \quad (14)$$

The resulting soft Complementary output is denoted by

$$\mathbf{y}_{\text{comp}}^{\text{soft}} = \text{ANFIS}_{\text{comp}}^{\text{soft}}(\mathbf{x}_{\text{comp}}; \tilde{\mathbf{g}}_{\text{comp}}), \quad (15)$$

where the superscript “soft” indicates that consequent-only soft gating is used. In this stage, variables already selected by the Primary hard mask are excluded from the Complementary antecedent through the fixed complementary-routing constraint. The soft Complementary training objective is

$$L_{S2}^{\text{soft}} = \text{Loss}(\mathbf{y}_{\text{primary}}^{\text{hard}} + \mathbf{y}_{\text{comp}}^{\text{soft}}; \mathbf{y}) + \lambda_{\text{comp}} \|\tilde{\mathbf{g}}_{\text{comp}}\|_1. \quad (16)$$

Here, the ℓ_1 penalty is imposed on the soft Complementary gate so that sparsity is enforced only on variables that remain available after Primary routing. In this context, residual correction means an additive logit or class-score correction optimized through the combined classification loss, not a separate regression model trained on explicit residual targets. The Complementary module is therefore evaluated through the combined prediction with the frozen Primary output, rather than as an independently strong standalone classifier.

After soft training, we binarize the learned Complementary gate using the threshold τ_{comp} to obtain

$$\mathbf{g}_{\text{comp}}^{\text{hard}} = \mathbb{I}[\mathbf{g}_{\text{comp}} \geq \tau_{\text{comp}}]. \quad (17)$$

The final Complementary hard mask is then defined as

$$\tilde{\mathbf{g}}_{\text{comp}}^{\text{hard}} = (\mathbf{1} - \mathbf{g}_{\text{primary}}^{\text{hard}}) \odot \mathbf{g}_{\text{comp}}^{\text{hard}}. \quad (18)$$

The resulting hard-routed Complementary output is denoted by

$$\mathbf{y}_{\text{comp}}^{\text{hard}} = \text{ANFIS}_{\text{comp}}^{\text{hard}}(\mathbf{x}_{\text{comp}}; \mathbf{g}_{\text{comp}}^{\text{hard}}), \quad (19)$$

where the superscript “hard” indicates that the fixed Complementary hard mask is applied to both the antecedent and consequent parts.

The hard Complementary refinement objective is

$$L_{S2}^{\text{hard}} = \text{Loss}(\mathbf{y}_{\text{primary}}^{\text{hard}} + \mathbf{y}_{\text{comp}}^{\text{hard}}, \mathbf{y}). \quad (20)$$

Because the hard Complementary mask is fixed after thresholding, no gate regularization term is used in this refinement stage. Therefore, Stage 2 hard refinement fine-tunes the Complementary ANFIS under the fixed complementary routing pattern determined by $\mathbf{g}_{\text{primary}}^{\text{hard}}$ and $\mathbf{g}_{\text{comp}}^{\text{hard}}$.

Figure 3 illustrates the complete Stage 2 procedure. The left panel shows consequent-only soft training of the Complementary module under the complementary-routing constraint, and the right panel shows thresholding of the learned Complementary gate followed by hard-routing refinement under the fixed Complementary hard mask.

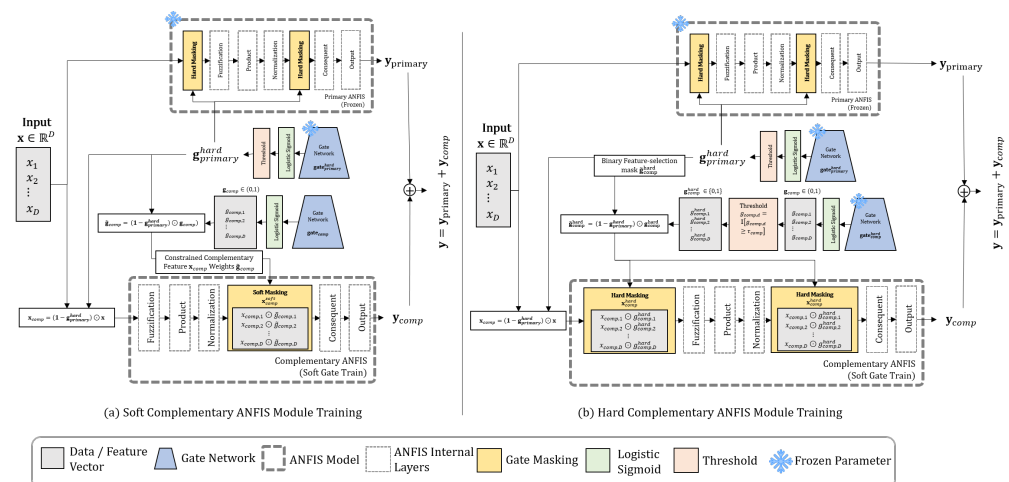


Figure 3. Stage 2 training workflow for the Complementary module. The Complementary module is trained after freezing the Primary module and is restricted to variables not selected by the Primary hard mask. The left panel shows soft Complementary training, and the right panel shows hard-routing refinement.

3.4.3. Rationale for Stage-Wise Optimization

1. **Enforced Specialization:** By freezing the Primary module and its hard routing during Stage 2, we prevent the Complementary module from simply copying the dominant decision function and encourage a distinct correction role.
2. **Sparsity Preservation:** The ℓ_1 penalty encourages compact soft gates, which are later converted into discrete module-specific variable subsets.
3. **Gradient Stability:** Consequent-only soft gating preserves stable antecedent membership computation during differentiable optimization.
4. **Interpretability:** Hard-gate masks make variable-role assignments explicit by showing which variables participate in the antecedent and consequent parts of each module.

3.5. Membership Function Configuration

3.5.1. Adaptive Membership Function Count

The number of membership functions per input can be fixed or determined adaptively. In this work, we use a fixed number n for all inputs to balance interpretability and expressiveness. Common choices are $n \in \{2, 3, 4\}$.

3.5.2. Membership Function Type

We employ Gaussian membership functions as the primary choice due to their smoothness properties, which improve gradient flow during backpropagation. This is consistent with the Gaussian MF definition introduced in Equation (21).

$$\mu_{A_{d,j}}(x_d) = \exp\left(-\frac{(x_d - c_{d,j})^2}{2\sigma_{d,j}^2}\right), \quad (21)$$

The membership function parameters (centers $c_{d,j}$ and widths $\sigma_{d,j}$) are initialized via grid partitioning: each input dimension is divided into n equal intervals, the center of each interval becomes the corresponding membership-function center, and the width is initialized as half of the interval width. In this study, ANFIS is introduced in a compact form centered on IF–THEN fuzzy rules. For a task with C output classes, a first-order TSK rule is written as

$$\text{Rule } r: \text{ IF } x_d \text{ is } A_{r,d} \ (d = 1, \dots, D), \text{ THEN } f_{r,c}(\mathbf{x}) = p_{r,c,0} + \sum_{d=1}^D p_{r,c,d} x_d, \quad c = 1, \dots, C. \quad (22)$$

Thus, for multi-class tasks, each rule has a vector consequent $\mathbf{f}_r(\mathbf{x}) = [f_{r,1}(\mathbf{x}), \dots, f_{r,C}(\mathbf{x})]^\top \in \mathbb{R}^C$. For binary tasks, we use $C = 1$ and omit the class index when this causes no ambiguity.

3.6. Overall Algorithm

Algorithm 1 summarizes the two-stage training procedure of GRS-ANFIS, including Primary soft training, Primary hard-routing refinement, Complementary soft training under the Primary–Complementary constraint, and Complementary hard-routing refinement.

Algorithm 1 Two-stage training procedure of GRS-ANFIS

Require: Training set $\mathcal{D} = \{(\mathbf{x}^{(i)}, \mathbf{y}^{(i)})\}_{i=1}^N$; sparsity coefficients $\lambda_{\text{primary}}, \lambda_{\text{comp}}$; thresholds $\tau_{\text{primary}}, \tau_{\text{comp}}$; epochs $T_{S1}, T_{S1}^{\text{hard}}, T_{S2}, T_{S2}^{\text{hard}}$

Ensure: Trained $\text{ANFIS}_{\text{primary}}, \text{ANFIS}_{\text{comp}}, \mathbf{g}_{\text{primary}}^{\text{hard}}$, and $\mathbf{g}_{\text{comp}}^{\text{hard}}$

1: Initialize $\text{ANFIS}_{\text{primary}}$ parameters

2: Initialize $\text{ANFIS}_{\text{comp}}$ parameters

3: Initialize gate logits $\mathbf{a}_{\text{primary}}$ and $\mathbf{a}_{\text{comp}}$

Primary soft call: $\text{ANFIS}_{\text{primary}}^{\text{soft}}(\mathbf{x}; \mathbf{g}_{\text{primary}})$ uses raw $\mathbf{x}$ in the antecedent and applies $\mathbf{g}_{\text{primary}}$ only to consequent coefficients.

Primary hard call: $\text{ANFIS}_{\text{primary}}^{\text{hard}}(\mathbf{x}; \mathbf{g}_{\text{primary}}^{\text{hard}})$ applies $\mathbf{g}_{\text{primary}}^{\text{hard}}$ to both antecedent and consequent.

Complementary soft call: $\text{ANFIS}_{\text{comp}}^{\text{soft}}(\mathbf{x}_{\text{comp}}; \mathbf{g}_{\text{primary}}^{\text{hard}}, \tilde{\mathbf{g}}_{\text{comp}})$ restricts the antecedent to the variables with $\mathbf{g}_{\text{primary}}^{\text{hard}} = 0$ and applies $\tilde{\mathbf{g}}_{\text{comp}}$ only to consequent coefficients.

Complementary hard call: $\text{ANFIS}_{\text{comp}}^{\text{hard}}(\mathbf{x}_{\text{comp}}; \mathbf{g}_{\text{comp}}^{\text{hard}})$ applies $\mathbf{g}_{\text{comp}}^{\text{hard}}$ to both antecedent and consequent.

Stage 1: Primary module training

4: **for** $t = 1$ to T_{S1} **do**

5: **for all** mini-batches $(\mathbf{x}, \mathbf{y})$ in $\mathcal{D}$ **do**

6: $\mathbf{g}_{\text{primary}} \leftarrow \text{sigmoid}(\mathbf{a}_{\text{primary}})$

7: $\mathbf{y}_{\text{primary}}^{\text{soft}} \leftarrow \text{ANFIS}_{\text{primary}}^{\text{soft}}(\mathbf{x}; \mathbf{g}_{\text{primary}})$

8: $\mathcal{L}_{S1} \leftarrow \text{Loss}(\mathbf{y}_{\text{primary}}^{\text{soft}}, \mathbf{y}) + \lambda_{\text{primary}} \|\mathbf{g}_{\text{primary}}\|_1$

9: Update $\text{ANFIS}_{\text{primary}}$ and $\mathbf{a}_{\text{primary}}$ by gradient descent on $\mathcal{L}_{S1}$

Algorithm 1 *Cont.*

```

10:   end for
11: end for
12:  $\mathbf{g}_{\text{primary}} \leftarrow \text{sigmoid}(\mathbf{a}_{\text{primary}})$ 
13:  $\mathbf{g}_{\text{primary}}^{\text{hard}} \leftarrow \mathbb{I}[\mathbf{g}_{\text{primary}} \geq \tau_{\text{primary}}]$ 
14: Freeze  $\mathbf{g}_{\text{primary}}^{\text{hard}}$ 
15: for  $t = 1$  to  $T_{S1}^{\text{hard}}$  do
16:   for all mini-batches  $(\mathbf{x}, \mathbf{y})$  in  $\mathcal{D}$  do
17:      $\mathbf{y}_{\text{primary}}^{\text{hard}} \leftarrow \text{ANFIS}_{\text{primary}}^{\text{hard}}(\mathbf{x}; \mathbf{g}_{\text{primary}}^{\text{hard}})$ 
18:      $\mathcal{L}_{S1}^{\text{hard}} \leftarrow \text{Loss}(\mathbf{y}_{\text{primary}}^{\text{hard}}, \mathbf{y})$ 
19:     Update only  $\text{ANFIS}_{\text{primary}}$  under fixed Primary gate masking
20:   end for
21: end for

Stage 2: Complementary module training
22: Freeze  $\text{ANFIS}_{\text{primary}}$  and  $\mathbf{g}_{\text{primary}}^{\text{hard}}$ 
23: for  $t = 1$  to  $T_{S2}$  do
24:   for all mini-batches  $(\mathbf{x}, \mathbf{y})$  in  $\mathcal{D}$  do
25:      $\mathbf{x}_{\text{comp}} \leftarrow (\mathbf{1} - \mathbf{g}_{\text{primary}}^{\text{hard}}) \odot \mathbf{x}$ 
26:      $\mathbf{g}_{\text{comp}} \leftarrow \text{sigmoid}(\mathbf{a}_{\text{comp}})$ 
27:      $\tilde{\mathbf{g}}_{\text{comp}} \leftarrow (\mathbf{1} - \mathbf{g}_{\text{primary}}^{\text{hard}}) \odot \mathbf{g}_{\text{comp}}$ 
28:      $\mathbf{y}_{\text{comp}} \leftarrow \text{ANFIS}_{\text{comp}}^{\text{soft}}(\mathbf{x}_{\text{comp}}, \tilde{\mathbf{g}}_{\text{comp}})$ 
29:      $\mathbf{y}_{\text{primary}} \leftarrow \text{ANFIS}_{\text{primary}}^{\text{hard}}(\mathbf{x}, \mathbf{g}_{\text{primary}}^{\text{hard}})$ 
30:      $\tilde{\mathbf{y}} \leftarrow \mathbf{y}_{\text{primary}} + \mathbf{y}_{\text{comp}}$ 
31:      $\mathcal{L}_{S2} \leftarrow \text{Loss}(\tilde{\mathbf{y}}, \mathbf{y}) + \lambda_{\text{comp}} \|\tilde{\mathbf{g}}_{\text{comp}}\|_1$ 
32:     Update  $\text{ANFIS}_{\text{comp}}$  and  $\mathbf{a}_{\text{comp}}$  by gradient descent on  $\mathcal{L}_{S2}$ 
33:   end for
34: end for
35:  $\mathbf{g}_{\text{comp}} \leftarrow \text{sigmoid}(\mathbf{a}_{\text{comp}})$ 
36:  $\mathbf{g}_{\text{comp}}^{\text{hard}} \leftarrow \mathbb{I}[\mathbf{g}_{\text{comp}} \geq \tau_{\text{comp}}]$ 
37:  $\tilde{\mathbf{g}}_{\text{comp}}^{\text{hard}} \leftarrow (\mathbf{1} - \mathbf{g}_{\text{primary}}^{\text{hard}}) \odot \mathbf{g}_{\text{comp}}^{\text{hard}}$ 
38: Freeze  $\tilde{\mathbf{g}}_{\text{comp}}^{\text{hard}}$ 
39: for  $t = 1$  to  $T_{S2}^{\text{hard}}$  do
40:   for all mini-batches  $(\mathbf{x}, \mathbf{y})$  in  $\mathcal{D}$  do
41:      $\mathbf{x}_{\text{comp}} \leftarrow (\mathbf{1} - \mathbf{g}_{\text{primary}}^{\text{hard}}) \odot \mathbf{x}$ 
42:      $\mathbf{y}_{\text{comp}} \leftarrow \text{ANFIS}_{\text{comp}}^{\text{hard}}(\mathbf{x}_{\text{comp}}, \tilde{\mathbf{g}}_{\text{comp}}^{\text{hard}})$ 
43:      $\mathbf{y}_{\text{primary}} \leftarrow \text{ANFIS}_{\text{primary}}^{\text{hard}}(\mathbf{x}, \mathbf{g}_{\text{primary}}^{\text{hard}})$ 
44:      $\tilde{\mathbf{y}} \leftarrow \mathbf{y}_{\text{primary}} + \mathbf{y}_{\text{comp}}$ 
45:      $\mathcal{L}_{S2}^{\text{hard}} \leftarrow \text{Loss}(\tilde{\mathbf{y}}, \mathbf{y})$ 
46:     Update only  $\text{ANFIS}_{\text{comp}}$  under fixed Complementary gate masking
47:   end for
48: end for

```

3.7. Inference and Interpretability**3.7.1. Forward Pass**

During inference, given an input $\mathbf{x}$, the Primary module uses the fixed hard mask $\mathbf{g}_{\text{primary}}^{\text{hard}}$ and the Complementary module uses the hard mask $\tilde{\mathbf{g}}_{\text{comp}}^{\text{hard}}$ obtained from the training stage.

As in all ANFIS-family models considered in this work, HTSK is used to compute antecedent rule scores, while the consequents remain masked linear functions.

The Primary rule scores are computed only on the selected antecedent variables:

$$z_{r,\text{primary}}(\mathbf{x}) = \frac{1}{\sum_{d=1}^D g_{\text{primary},d}^{\text{hard}} + \epsilon} \sum_{d=1}^D g_{\text{primary},d}^{\text{hard}} \log(\mu_{A_{r,d}^P}(x_d) + \epsilon). \quad (23)$$

The normalized Primary firing strength is

$$\bar{w}_{r,\text{primary}}(\mathbf{x}) = \frac{\exp(z_{r,\text{primary}}(\mathbf{x}))}{\sum_j \exp(z_{j,\text{primary}}(\mathbf{x}))}. \quad (24)$$

The Primary consequent is

$$f_{r,\text{primary}}(\mathbf{x}) = p_{r,0}^{\text{primary}} + \sum_{d=1}^D g_{\text{primary},d}^{\text{hard}} p_{r,d}^{\text{primary}} x_d. \quad (25)$$

The Primary output is

$$\mathbf{y}_{\text{primary}} = \sum_r \bar{w}_{r,\text{primary}}(\mathbf{x}) f_{r,\text{primary}}(\mathbf{x}). \quad (26)$$

The Complementary rule scores are computed on the variables selected by $\tilde{g}_{\text{comp}}^{\text{hard}}$:

$$z_{r,\text{comp}}(\mathbf{x}) = \frac{1}{\sum_{d=1}^D \tilde{g}_{\text{comp},d}^{\text{hard}} + \epsilon} \sum_{d=1}^D \tilde{g}_{\text{comp},d}^{\text{hard}} \log(\mu_{A_{r,d}^c}(x_d) + \epsilon). \quad (27)$$

The normalized Complementary firing strength is

$$\bar{w}_{r,\text{comp}}(\mathbf{x}) = \frac{\exp(z_{r,\text{comp}}(\mathbf{x}))}{\sum_j \exp(z_{j,\text{comp}}(\mathbf{x}))}. \quad (28)$$

The Complementary consequent is

$$f_{r,\text{comp}}(\mathbf{x}) = p_{r,0}^{\text{comp}} + \sum_{d=1}^D \tilde{g}_{\text{comp},d}^{\text{hard}} p_{r,d}^{\text{comp}} x_d. \quad (29)$$

The Complementary output is

$$\mathbf{y}_{\text{comp}} = \sum_r \bar{w}_{r,\text{comp}}(\mathbf{x}) f_{r,\text{comp}}(\mathbf{x}). \quad (30)$$

Finally, the two module outputs are combined into the summed module score

$$\tilde{\mathbf{y}} = \mathbf{y}_{\text{primary}} + \mathbf{y}_{\text{comp}}. \quad (31)$$

For classification, the summed module score is converted to the final class probability by a sigmoid (binary) or softmax (multi-class) function, and the predicted label is derived from that probability.

3.7.2. Dual-Layer Interpretability

GRS-ANFIS provides interpretability at two levels:

Level 1—Feature Routing.

The binary Primary mask $\mathbf{g}_{\text{primary}}^{\text{hard}}$ and the binary Complementary mask $\tilde{\mathbf{g}}_{\text{comp}}^{\text{hard}}$ explicitly indicate which variables participate in the Primary module and which remaining variables are reserved for Complementary correction.

Level 2—Role-Specific Fuzzy Rules.

For each module $b \in \{\text{Primary}, \text{Complementary}\}$, the fuzzy rules are defined on the hard-gated subset S_b :

$$\text{Rule } r : \text{ IF } x_d \text{ is } A_{r,d}^b \text{ for } d \in S_b, \text{ THEN } \mathbf{y} = f_r(\mathbf{x}; S_b). \quad (32)$$

Here, the antecedent firing and consequent evaluation both depend only on the variables in S_b .

Together, these two levels enable a human expert to inspect (1) which variables are assigned to each role, (2) which module-specific fuzzy rules are activated, and (3) how each activated rule contributes through its local linear consequent.

4. Experiments

4.1. Experimental Setup

Datasets

We evaluate GRS-ANFIS on four public tabular classification benchmarks spanning low-to-high dimensional settings. ANFIS, H-ANFIS, GRS-ANFIS, SVM, and the tree-based and additive tabular baselines (Random Forest, XGBoost, LightGBM, CatBoost, and EBM) are evaluated using the full preprocessed input representation for each dataset, whereas GA-ANFIS and PSO-ANFIS additionally incorporate feature-selection effects within their optimization pipelines. For Breast Cancer and Vowel, the source datasets contain 9 and 27 raw variables, respectively, and their one-hot/preprocessed model-input representations contain 80 and 29 dimensions, respectively. These preprocessed dimensions are also the basis for selected-feature counts in the supplemental gate-sensitivity analyses. Table 1 summarizes the benchmark datasets used in the evaluation.

Table 1. Benchmark datasets used in evaluation.

Dataset	Task	Raw Feat.	Input Dim.	Notes
Breast Cancer [17]	Binary classification	9	80	9 raw diagnostic variables; 80 one-hot/preprocessed model-input dimensions
Vowel [18]	Multi-class classification	27	29	Vowel recognition (11 classes); 29 one-hot/preprocessed model-input dimensions
Spambase [19]	Binary classification	57	57	UCI email spam detection
Gisette [20]	Binary classification	5000	5000	High-dimensional feature setting

In Appendix A, selected-feature counts refer to the encoded/preprocessed input dimensions used by the gate masks. Thus, Breast Cancer selected counts such as 77.6, 70.4, and 51.2 in the sparsity diagnostic are counts out of the 80 one-hot/preprocessed input dimensions, not counts out of the 9 raw diagnostic variables. Likewise, Vowel selected counts such as 29.0, 28.6, and 21.2 are counts out of the 29 one-hot/preprocessed input dimensions, not counts out of the 27 raw variables.

4.2. Hyperparameter Selection

Key hyperparameters and their roles are summarized in Table 2.

Table 2. Hyperparameter setting.

Hyperparameter	Role	Typical Range
$\lambda_{\text{primary}}, \lambda_{\text{comp}}$	ℓ_1 sparsity penalty	10^{-4} to 10^{-1}
$\tau_{\text{primary}}, \tau_{\text{comp}}$	Hard gate thresholds	0.3 to 0.7 (usually 0.5)
n	Membership functions per input	2 to 5
Learning rate η	Optimizer step size	10^{-4} to 10^{-2}
$\text{max_epochs}_{S1, S2}$	Stage-wise maximum epochs	10 to 100

All hard-gate thresholds are set to 0.5. This threshold is applied to sigmoid gate values and is equivalent to thresholding the corresponding gate logits at zero.

Preprocessing and protocol:

- Features are standardized using training-fold statistics and then applied to the corresponding validation fold.
- We report mean $\pm$ standard deviation over 5 stratified folds.

4.2.1. Baseline Models

We compare GRS-ANFIS against representative neuro-fuzzy baselines, a reference non-fuzzy classifier, and tree-based and additive tabular baselines:

1. ANFIS: Standard TSK-ANFIS with a fixed rule budget (Rule = 30). For computational feasibility in high-dimensional settings, the implemented ANFIS baseline uses a compact shared-MF parameterization with a fixed rule budget, rather than the full Cartesian grid-partition rule construction. Specifically, Gaussian membership functions are shared across rules within each input dimension.
2. GA-ANFIS: ANFIS optimized with a genetic algorithm, including feature-selection effects.
3. PSO-ANFIS: ANFIS optimized with particle swarm optimization, including feature-selection effects.
4. H-ANFIS: Parallel hierarchical ANFIS baseline with average fusion.
5. SVM: Support Vector Machine with RBF kernel [21].
6. GRS-ANFIS (Primary): Primary module only.
7. GRS-ANFIS (Full): Primary + Complementary modules.
8. Random Forest: Tree-ensemble classifier used as a strong tabular baseline [22].
9. XGBoost: Gradient-boosted tree baseline reported with default model hyperparameters [23].
10. LightGBM: Supplemental gradient-boosting baseline [24].
11. CatBoost: Supplemental gradient-boosting baseline [25].
12. Explainable Boosting Machine (EBM): Intrinsically interpretable generalized additive model baseline [26,27].

Random Forest and XGBoost are reported as default-hyperparameter baselines in the dataset-specific performance tables. LightGBM, CatBoost, and EBM are evaluated as supplemental tabular baselines; their Accuracy, Precision, Recall, and F1 values are reported in the same performance tables and summarized in the Appendix A.

4.2.2. Hyperparameter Settings

Common protocol.

The ANFIS-family and SVM hyperparameters are selected under the same fold-wise protocol; that is, they are tuned/selected within each training fold using the corresponding validation split. Random Forest and XGBoost use library default model hyperparameters, with only reproducibility and execution-control options set. LightGBM, CatBoost, and EBM are evaluated in supplemental baseline analyses. The configurations reported below, therefore, correspond to the stated protocol for each baseline family, not the exhaustive best-case optima for every model.

GRS-ANFIS training.

GRS-ANFIS is trained in two stages for all datasets: Stage 1 trains the Primary module only; Stage 2 trains the Complementary module for correction while freezing the Primary. After each stage, hard-routing refinement is applied. Unless stated otherwise, all mask thresholds are set to $\tau = 0.5$ and hard-stage sparsity penalties are set to 0. The common experimental settings are summarized in Table 3.

Table 3. Common experimental settings used across all datasets.

Item	Setting
ANFIS (baseline)	Learning rate $\eta = 0.1$, number of rules $R = 30$ (for a standard TSK-ANFIS only), epochs = 100, selected_features = $\emptyset$, Membership Functions per input = 3 (default).
SVM (RBF)	$C = 1.0$, $\gamma = \text{scale}$, kernel = rbf, probability = True, random_state = 42.
Genetic Algorithm (GA)	population = 20, generations = 10. Roulette-wheel selection probability is computed per generation: $\tilde{f}_i = f_i - \min_j f_j$, $p_i = \tilde{f}_i / \sum_j \tilde{f}_j$; if $\sum_j \tilde{f}_j \leq 0$, use uniform $p_i = 1/N$. Mutation rate = mutation_percent/100 = 0.10.
Particle Swarm Optimization (PSO)	particles = 20, iterations = 10.
H-ANFIS split rule	Random half split. Given feature index order $[0, \dots, d-1]$, shuffle by split_seed; cut = $\max(1, \lfloor d/2 \rfloor)$; G_a = first cut indices, G_b = remaining indices. If $d < 2$, an error is raised.
H-ANFIS split seed	split_seed = 42 + offset + 997 · fold_idx. For H-ANFIS, we use offset = 101 (stacked variant excluded).

Dataset-specific hyperparameter settings are reported in Tables 4–6.

Table 4. Dataset-specific hyperparameters for GA-ANFIS and PSO-ANFIS.

Dataset	GA-ANFIS				PSO-ANFIS			
	LR	Rules	Epochs	#Sel	LR	Rules	Epochs	#Sel
BCWD	0.05	29	47	5	0.10	20	16	7
Vowel	0.05	26	47	23	0.10	4	61	19
Spambase	0.05	25	51	39	0.10	12	40	31
Gisette	0.05	9	25	91	0.01	25	48	87

LR: learning rate; Rules: number of fuzzy rules; Epochs: training epochs; #Sel: number of selected features. Fixed settings: GA uses population = 20, generations = 10, and mutation rate = 0.10 (i.e., mutation_percent = 10). Selection is roulette-wheel based on shifted fitness: $\tilde{f}_i = f_i - \min_j f_j$, $p_i = \tilde{f}_i / \sum_j \tilde{f}_j$; if $\sum_j \tilde{f}_j \leq 0$, use uniform $p_i = 1/N$. PSO uses particles = 20 and iterations = 10.

Table 5. Dataset-specific hyperparameters for H-ANFIS.

Dataset	LR	Epochs	Module Rules	Top Rules	MF	wd
BCWD	0.05	80	16	8	2	10^{-5}
Vowel	0.05	80	18	8	2	10^{-5}
Spambase	0.05	80	20	8	2	10^{-5}
Gisette	0.03	70	14	6	2	10^{-5}

Table 6. Dataset-specific hyperparameters for GRS-ANFIS.

Dataset	LR (P/C)	Rules (P/C)	MF	Epochs (S1/S2)	λ (P/C)	wd	Hard Epochs (P/C)
BCWD	0.01/0.005	7/11	2	80/10	0.01/0.0	10^{-5}	5/5
Vowel	0.1/0.005	7/11	2	80/10	0.01/0.0	10^{-5}	30/10
Spambase	0.1/0.01	7/11	3	60/50	0.001/0.001	10^{-6}	50/50
Gisette	0.01/0.01	7/11	2	30/30	0.01/0.001	10^{-4}	20/10

LR: learning rate; P/C: Primary/Complementary; MF: membership functions per feature; Epochs (S1/S2): training epochs for Stage 1 (Primary only)/Stage 2 (Complementary with frozen Primary); λ : sparsity coefficient; wd: weight decay; Hard epochs: epochs for hard-routing refinement. Fixed settings: all mask thresholds are set to $\tau = 0.5$; hard-stage penalty is not used (set to 0).

4.2.3. Evaluation Metrics

We evaluate both predictive performance and rule-based interpretability.

(A) Predictive performance metrics.

We report Accuracy and F1 score. For binary tasks, F1 is the standard harmonic mean of Precision and Recall. For multi-class tasks, we report the support-weighted average of class-wise F1 scores.

$$\text{Acc} = \frac{\#\{\hat{\mathbf{y}}_i = \mathbf{y}_i\}}{n}, \quad (33)$$

$$\text{F1}_{\text{binary}} = \frac{2 \text{Precision} \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (34)$$

and for multi-class tasks

$$\text{F1}_{\text{weighted}} = \sum_{c=1}^C \frac{n_c}{n} \text{F1}_c, \quad \text{F1}_c = \frac{2 \text{Precision}_c \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c}, \quad (35)$$

where n_c is the number of test samples in class c .

(B) Interpretability metrics.

(B1) Nauck index for ANFIS. The Nauck index summarizes interpretability as a product of three human-readable factors: (i) complexity (fewer rules/antecedents are easier to read), (ii) coverage (partitions should adequately cover the domain), (iii) partition granularity (too many membership functions harms readability). Following Nauck [28],

$$I_{\text{Nauck}} = I_{\text{complex}} \cdot I_{\text{cover}} \cdot I_{\text{partition}}. \quad (36)$$

Complexity (I_{complex}). This term decreases as the number of rules and antecedent variables increases; thus, a larger I_{complex} indicates a structurally simpler rule base. Let r be the number of rules, n_k the number of antecedent variables used in rule k , and m the number of classes.

$$I_{\text{complex}} = \frac{m}{\sum_{k=1}^r n_k}. \quad (37)$$

Coverage (I_{cover}). For each input variable x_i with domain X_i and p_i membership functions $\{\mu_i^{(1)}, \dots, \mu_i^{(p_i)}\}$, define

$$h_i(\mathbf{x}) = \sum_{k=1}^{p_i} \mu_i^{(k)}(\mathbf{x}), \quad (38)$$

and

$$\tilde{h}_i(\mathbf{x}) = \begin{cases} h_i(\mathbf{x}), & 0 < h_i(\mathbf{x}) < 1, \\ \frac{p_i - h_i(\mathbf{x})}{p_i - 1}, & \text{otherwise.} \end{cases} \quad (39)$$

Then the coverage of variable i is

$$\text{cover}_i = \frac{1}{N_i} \int_{X_i} \tilde{h}_i(x) dx, \quad N_i = \int_{X_i} dx, \quad (40)$$

where the integral can be replaced by a sum for discrete finite domains. Finally, we average across used variables:

$$I_{\text{cover}} = \frac{1}{d} \sum_{i=1}^d \text{cover}_i, \quad (41)$$

with d being the number of input variables appearing in the rule base.

Partition ($I_{\text{partition}}$). The partition index penalizes too many membership functions:

$$\text{partition}_i = \frac{1}{p_i - 1}, \quad I_{\text{partition}} = \frac{1}{d} \sum_{i=1}^d \text{partition}_i. \quad (42)$$

A larger $I_{\text{Nauck}} \in [0, 1]$ indicates a more interpretable (simpler and better-structured) fuzzy rule base.

(B2) Unified rule-based interpretability index. To report interpretability in a single comparative table, we use a unified Nauck-based interpretability index. For ANFIS, GA-ANFIS, and PSO-ANFIS, the reported value is the Nauck index itself. For H-ANFIS and GRS-ANFIS, the reported value is HFSi [29], where the subsystem-level interpretability term $E_{ij} \in [0, 1]$ is instantiated by the Nauck index computed from the corresponding subsystem rule base.

Let the hierarchical or modular model have n layers, where layer i contains m_i subsystems. Then HFSi is defined as

$$\text{HFSi} = \sum_{i=1}^n l_i \left(\frac{1}{m_i} \sum_{j=1}^{m_i} E_{ij} \right), \quad (43)$$

with layer weights

$$l_i = \frac{2(n-i+1)}{n(n+1)}, \quad i = 1, \dots, n. \quad (44)$$

In this study, we set

$$E_{ij} = I_{\text{Nauck}}$$

for each subsystem. Therefore, HFSi preserves the Nauck-based interpretability meaning at the subsystem level while additionally accounting for hierarchical or modular structure through layer-weighted aggregation.

Accordingly, all interpretability values reported in this paper are presented on a unified Nauck-based scale: they are equal to the Nauck index for ANFIS, GA-ANFIS, and PSO-ANFIS, and equal to HFSi for H-ANFIS and GRS-ANFIS.

(C) Complementary correction metrics (Q1 boundary analysis).

To quantify where the Complementary module helps most, we compute confidence bins from Primary predictions within each fold:

$$\text{conf}(\mathbf{x}) = \begin{cases} \max(\mathbf{p}_{\text{primary}}, 1 - \mathbf{p}_{\text{primary}}), & \text{binary,} \\ \max_c \mathbf{p}_{\text{primary}}(c), & \text{multi-class.} \end{cases} \quad (45)$$

Samples are split into equal-frequency bins ($n_{\text{bins}} = 4$). Specifically, for a fold with N validation samples, let r_i denote the ascending rank of the base confidence of sample $\mathbf{x}_i$ within that fold. Then the bin assignment is

$$\text{bin}_i = \left\lceil \frac{r_i}{N} \cdot 4 \right\rceil. \quad (46)$$

Q1 denotes the lowest-confidence quartile, i.e.

$$\mathbf{x}_i \in \text{Q1} \iff \text{bin}_i = 1. \quad (47)$$

Thus, Q1 contains the lowest confidence, approximately 25% of validation samples in each fold under rank-based equal-frequency binning. Note that quantile values, if recorded,

are used only for logging; actual Q1 membership is determined by the rank-based binning above. For Q1 we report:

$$\Delta\text{Acc}_{\text{Q1}} = \text{Acc}_{\text{Full,Q1}} - \text{Acc}_{\text{Primary,Q1}}, \quad (48)$$

wrong $\rightarrow$ right and right $\rightarrow$ wrong flips, and Error Recovery Rate:

$$\text{Error Recovery Rate} = \frac{\#(w \rightarrow r)}{\#(w \rightarrow r) + \#(\text{stable wrong})}. \quad (49)$$

Here, stable wrong denotes samples misclassified by the Primary module and still misclassified after adding the Complementary module.

4.2.4. Cross-Validation Protocol

We use stratified 5-fold cross-validation with shuffled splits and a fixed random seed. Final metrics are reported as mean $\pm$ standard deviation across the five folds. The Q1 boundary analysis is reported as a descriptive comparison between Primary-only and Full predictions in the lowest-confidence region.

5. Results

5.1. Predictive Performance and Interpretability

Tables 7–10 summarize predictive performance (Accuracy and F1) together with a unified Nauck-based interpretability index. For ANFIS, GA-ANFIS, and PSO-ANFIS, the reported interpretability value is the Nauck index itself. For H-ANFIS and GRS-ANFIS, the reported value is HFSi computed from subsystem-level Nauck indices. In this way, all compared model families are reported on a common rule-based interpretability scale, while hierarchical or modular structure is additionally reflected when applicable. In the four main performance tables, non-ANFIS tabular baselines are listed first, and ANFIS-family models are grouped below them so that overall tabular competitiveness and within-family neuro-fuzzy comparisons can be read separately.

Table 7. Vowel (11 classes): performance and interpretability (mean $\pm$ std). Rows are grouped with non-ANFIS tabular baselines first and ANFIS-family models below them. For each numeric metric, the best value across all rows is shown in bold and the second-best value is underlined; for n_{rules} , fewer rules are treated as more compact. Ties are marked jointly. The reported interpretability value equals the Nauck index for ANFIS, GA-ANFIS, and PSO-ANFIS, and equals HFSi for H-ANFIS and GRS-ANFIS, where HFSi is computed from subsystem-level Nauck indices. Random Forest and XGBoost use library default hyperparameters and report Accuracy, Precision, Recall, and F1; LightGBM, CatBoost, and EBM report the same metrics from supplemental baseline analyses.

Model	Acc	Precision	Recall	F1	Nauck-Based Index	n_{rules}
Non-ANFIS tabular baselines						
SVM	0.8566 $\pm$ 0.0228	0.8679 $\pm$ 0.0193	0.8566 $\pm$ 0.0228	0.8544 $\pm$ 0.0237	–	–
Random Forest	0.9717 $\pm$ 0.0092	<u>0.9743</u> $\pm$ 0.0075	0.9717 $\pm$ 0.0092	0.9715 $\pm$ 0.0094	–	–
XGBoost	0.9131 $\pm$ 0.0308	0.9188 $\pm$ 0.0299	0.9131 $\pm$ 0.0308	0.9125 $\pm$ 0.0306	–	–
LightGBM	0.9475 $\pm$ 0.0216	0.9499 $\pm$ 0.0219	0.9475 $\pm$ 0.0216	0.9472 $\pm$ 0.0216	–	–
CatBoost	0.9071 $\pm$ 0.0181	0.9097 $\pm$ 0.0187	0.9071 $\pm$ 0.0181	0.9063 $\pm$ 0.0183	–	–
EBM	0.9091 $\pm$ 0.0211	0.9117 $\pm$ 0.0209	0.9091 $\pm$ 0.0211	0.9078 $\pm$ 0.0218	–	–

Table 7. Cont.

Model	Acc	Precision	Recall	F1	Nauck-Based Index	n_{rules}
ANFIS-family models						
ANFIS	0.9313 ± 0.0287	0.9355 ± 0.0259	0.9313 ± 0.0287	0.9313 ± 0.0288	0.0030 ± 0.0001	30
GA-ANFIS	0.7677 ± 0.0357	0.7698 ± 0.0351	0.7677 ± 0.0357	0.7641 ± 0.0348	0.0038 ± 0.0001	26
PSO-ANFIS	0.8293 ± 0.0356	0.8345 ± 0.0362	0.8293 ± 0.0356	0.8274 ± 0.0365	0.0434 ± 0.0001	4
H-ANFIS	0.9313 ± 0.0136	0.9356 ± 0.0129	0.9313 ± 0.0136	0.9310 ± 0.0140	0.0129 ± 0.0011	36
GRS-ANFIS(primary)	<u>0.9727</u> ± 0.0092	0.9743 ± 0.0082	<u>0.9727</u> ± 0.0092	<u>0.9725</u> ± 0.0094	<u>0.0253</u> ± 0.0019	<u>7</u>
GRS-ANFIS(full)	0.9737 ± 0.0055	0.9752 ± 0.0052	0.9737 ± 0.0055	0.9736 ± 0.0057	0.0127 ± 0.0010	18

Table 8. Spambase: performance and interpretability (mean ± std). Rows are grouped with non-ANFIS tabular baselines first and ANFIS-family models below them. For each numeric metric, the best value across all rows is shown in bold and the second-best value is underlined; for n_{rules} , fewer rules are treated as more compact. Ties are marked jointly. Random Forest and XGBoost use library default hyperparameters and report Accuracy, Precision, Recall, and F1; LightGBM, CatBoost, and EBM report the same metrics from supplemental baseline analyses.

Model	Acc	Precision	Recall	F1	Nauck-Based Index	n_{rules}
Non-ANFIS tabular baselines						
SVM	0.7118 ± 0.0125	0.7158 ± 0.0405	0.4490 ± 0.0156	0.5512 ± 0.0117	–	–
Random Forest	<u>0.9520</u> ± 0.0046	0.9509 ± 0.0102	0.9261 ± 0.0153	<u>0.9382</u> ± 0.0063	–	–
XGBoost	0.9509 ± 0.0075	0.9456 ± 0.0087	0.9288 ± 0.0153	0.9371 ± 0.0099	–	–
LightGBM	0.9550 ± 0.0066	0.9428 ± 0.0088	0.9432 ± 0.0169	0.9429 ± 0.0087	–	–
CatBoost	0.9511 ± 0.0057	<u>0.9467</u> ± 0.0071	0.9283 ± 0.0166	0.9373 ± 0.0079	–	–
EBM	0.9509 ± 0.0068	0.9426 ± 0.0064	<u>0.9321</u> ± 0.0202	0.9372 ± 0.0094	–	–
ANFIS-family models						
ANFIS	0.8596 ± 0.0102	0.9429 ± 0.0191	0.6861 ± 0.0410	0.7933 ± 0.0223	0.0000 ± 0.0000	30
GA-ANFIS	0.9096 ± 0.0078	0.9310 ± 0.0049	0.8323 ± 0.0243	0.8787 ± 0.0123	0.0000 ± 0.0000	25
PSO-ANFIS	0.8855 ± 0.0091	0.9216 ± 0.0149	0.7755 ± 0.0204	0.8421 ± 0.0136	0.0001 ± 0.0000	<u>12</u>
H-ANFIS	0.9135 ± 0.0033	0.9263 ± 0.0131	0.8483 ± 0.0158	0.8854 ± 0.0050	0.0002 ± 0.0000	40
GRS-ANFIS(primary)	0.9272 ± 0.0072	0.9339 ± 0.0088	0.8775 ± 0.0233	0.9046 ± 0.0108	<u>0.0005</u> ± 0.0000	7
GRS-ANFIS(full)	0.9315 ± 0.0034	0.9328 ± 0.0120	0.8908 ± 0.0202	0.9111 ± 0.0058	0.0006 ± 0.0001	18

Table 9. Breast Cancer: performance and interpretability (mean ± std). Rows are grouped with non-ANFIS tabular baselines first and ANFIS-family models below them. For each numeric metric, the best value across all rows is shown in bold and the second-best value is underlined; for n_{rules} , fewer rules are treated as more compact. Ties are marked jointly. Random Forest and XGBoost use library default hyperparameters and report Accuracy, Precision, Recall, and F1; LightGBM, CatBoost, and EBM report the same metrics from supplemental baseline analyses.

Model	Acc	Precision	Recall	F1	Nauck-Based Index	n_{rules}
Non-ANFIS tabular baselines						
SVM	0.9743 ± 0.0130	0.9414 ± 0.0297	0.9877 ± 0.0183	0.9637 ± 0.0181	–	–
Random Forest	0.9614 ± 0.0081	<u>0.9500</u> ± 0.0228	0.9378 ± 0.0006	0.9437 ± 0.0113	–	–
XGBoost	0.9542 ± 0.0166	0.9488 ± 0.0233	0.9172 ± 0.0383	0.9323 ± 0.0249	–	–
LightGBM	0.9571 ± 0.0113	0.9454 ± 0.0177	0.9296 ± 0.0310	0.9371 ± 0.0168	–	–
CatBoost	0.9599 ± 0.0108	0.9499 ± 0.0230	0.9335 ± 0.0274	0.9413 ± 0.0163	–	–
EBM	0.9671 ± 0.0082	0.9519 ± 0.0356	0.9544 ± 0.0226	0.9525 ± 0.0113	–	–
ANFIS-family models						
ANFIS	0.9471 ± 0.0108	0.9230 ± 0.0337	0.9254 ± 0.0311	0.9235 ± 0.0148	0.0001 ± 0.0000	30
GA-ANFIS	0.9485 ± 0.0061	0.9255 ± 0.0091	0.9253 ± 0.0236	0.9252 ± 0.0093	0.0001 ± 0.0000	29
PSO-ANFIS	0.9585 ± 0.0146	0.9351 ± 0.0242	0.9463 ± 0.0447	0.9400 ± 0.0218	0.0001 ± 0.0000	20
H-ANFIS	0.9599 ± 0.0063	0.9351 ± 0.0152	0.9503 ± 0.0232	0.9424 ± 0.0092	0.0004 ± 0.0000	32
GRS-ANFIS(primary)	0.9657 ± 0.0060	0.9371 ± 0.0304	0.9668 ± 0.0315	0.9510 ± 0.0086	<u>0.0009</u> ± 0.0000	7
GRS-ANFIS(full)	<u>0.9685</u> ± 0.0081	0.9407 ± 0.0296	<u>0.9710</u> ± 0.0112	<u>0.9553</u> ± 0.0108	0.0033 ± 0.0005	<u>18</u>

Table 10. Gisette: performance and interpretability (mean $\pm$ std) in the original 5000-feature setting. Rows are grouped with non-ANFIS tabular baselines first and ANFIS-family models below them. For each numeric metric, the best value across all rows is shown in bold and the second-best value is underlined; for n_{rules} , fewer rules are treated as more compact. Ties are marked jointly. Random Forest and XGBoost use library default hyperparameters and report Accuracy, Precision, Recall, and F1; LightGBM, CatBoost, and EBM report the same metrics from supplemental baseline analyses.

Model	Acc	Precision	Recall	F1	Nauck-Based Index	n_{rules}
Non-ANFIS tabular baselines						
SVM	0.9755 $\pm$ 0.0071	0.9716 $\pm$ 0.0095	0.9797 $\pm$ 0.0051	0.9756 $\pm$ 0.0070	–	–
Random Forest	0.9722 $\pm$ 0.0066	0.9830 $\pm$ 0.0103	0.9610 $\pm$ 0.0035	0.9719 $\pm$ 0.0065	–	–
XGBoost	0.9757 $\pm$ 0.0061	0.9767 $\pm$ 0.0085	0.9747 $\pm$ 0.0058	0.9757 $\pm$ 0.0060	–	–
LightGBM	0.9790 $\pm$ 0.0037	<u>0.9793</u> $\pm$ 0.0061	<u>0.9787</u> $\pm$ 0.0018	0.9790 $\pm$ 0.0037	–	–
CatBoost	0.9725 $\pm$ 0.0061	0.9721 $\pm$ 0.0095	0.9730 $\pm$ 0.0043	0.9725 $\pm$ 0.0060	–	–
EBM	0.9738 $\pm$ 0.0050	0.9737 $\pm$ 0.0084	0.9740 $\pm$ 0.0056	0.9738 $\pm$ 0.0049	–	–
ANFIS-family models						
ANFIS	0.9657 $\pm$ 0.0041	0.9726 $\pm$ 0.0046	0.9583 $\pm$ 0.0070	0.9654 $\pm$ 0.0041	0.0000 $\pm$ 0.0000	30
GA-ANFIS	0.9297 $\pm$ 0.0127	0.9288 $\pm$ 0.0193	0.9310 $\pm$ 0.0122	0.9298 $\pm$ 0.0122	0.0002 $\pm$ 0.0000	<u>9</u>
PSO-ANFIS	0.9257 $\pm$ 0.0068	0.9247 $\pm$ 0.0131	0.9270 $\pm$ 0.0116	0.9258 $\pm$ 0.0066	<u>0.0001</u> $\pm$ 0.0000	25
H-ANFIS	0.9713 $\pm$ 0.0044	0.9748 $\pm$ 0.0058	0.9677 $\pm$ 0.0040	0.9712 $\pm$ 0.0044	0.0000 $\pm$ 0.0000	28
GRS-ANFIS(primary)	0.9755 $\pm$ 0.0050	0.9760 $\pm$ 0.0087	0.9750 $\pm$ 0.0020	0.9755 $\pm$ 0.0049	0.0000 $\pm$ 0.0000	7
GRS-ANFIS(full)	<u>0.9758</u> $\pm$ 0.0048	0.9760 $\pm$ 0.0078	0.9757 $\pm$ 0.0025	<u>0.9758</u> $\pm$ 0.0047	0.0001 $\pm$ 0.0000	18

Overall, GRS-ANFIS achieves competitive accuracy while maintaining compact, structured rule bases via role-separated gating. Across datasets, the Full model (Primary + Complementary) tends to match or exceed the Primary-only model in Accuracy/F1, indicating that the Complementary module can add predictive value beyond the dominant Primary decision. In terms of interpretability, GRS-ANFIS typically uses substantially fewer rules than ANFIS baselines with a fixed rule budget (e.g., 30 rules), while gains in the unified Nauck-based interpretability index remain dataset-dependent and feature-level compactness remains data-dependent.

5.2. Computational Cost and Runtime Results

After the main predictive and interpretability results, we report the computational cost of the stage-wise routing procedure as a Results-section runtime analysis. We measured approximate wall-clock training time on a single representative fold under the same split policy used in the main experiments. Dataset sizes and input dimensions are reported separately in Table 1; Table 11 reports only timing quantities to avoid conflating data scale with computational cost. These timings are intended to characterize cost rather than predictive quality. The full GRS-ANFIS path includes Primary soft training, Primary hard-route refinement, Complementary residual training with the Primary route frozen, and Complementary hard-route refinement. Therefore, it is consistently more expensive than the Primary-only variant and substantially more expensive than Random Forest or LightGBM under the reported fixed comparator settings. Relative to standard ANFIS, the measured ratio depends on the fixed rule and epoch budgets; consequently, we interpret runtime as a transparent trade-off for role-separated decomposition rather than as an efficiency advantage.

Table 11. Approximate wall-clock training-time comparison on a representative fold. Times are seconds for final model fitting only. ANFIS and GRS-ANFIS use fixed manuscript/revision hyperparameters, while Random Forest and LightGBM use fixed comparator settings without additional hyperparameter search. Gisette ANFIS and GRS-ANFIS entries are per-epoch extrapolations from bounded probes.

Dataset	ANFIS	GRS Primary-Only	GRS Full	RF	LightGBM
Breast Cancer	2.88	0.93	1.87	0.22	0.04
Vowel	1.22	1.85	3.71	0.46	1.09
Spambase	11.54	4.76	12.00	1.05	0.35
Gisette	1570.97 *	208.43 *	522.40 *	11.38	34.47

* Per-epoch extrapolation from bounded probes rather than direct full training.

5.3. Complementary Module Analysis

To identify when the Complementary module is most useful, we evaluate boundary-near samples in the lowest-confidence quartile (Q1). Q1 is defined by equal-frequency binning of the Primary confidence within each fold. Table 12 reports Q1 accuracy for Primary-only and Full inference, wrong $\rightarrow$ right and right $\rightarrow$ wrong flips, and the Error Recovery Rate. The results show dataset-dependent correction effects in low-confidence regions. Spambase has the largest net gain (+1.74 pp), with 41 corrected boundary errors and 21 newly introduced errors. BCWD and Vowel also show positive Error Recovery Rates (17.6% and 12.0%). Gisette shows only marginal improvement, indicating limited complementary benefit under this setting. We do not report the Complementary branch as a standalone model in the main performance tables because it is not trained or used as an autonomous classifier. After the Primary module is fixed, the Complementary module is optimized under the Full-model classification loss using only the remaining routed variables. It therefore lacks the complete feature context required for autonomous classification and should be read as an auxiliary correction component. Two results support this view. First, the F1 columns of the main performance tables show small positive gains from Primary-only to Full inference: Vowel 0.9725 to 0.9736 (Table 7), Spambase 0.9046 to 0.9111 (Table 8), Breast Cancer 0.9510 to 0.9553 (Table 9), and Gisette 0.9755 to 0.9758 (Table 10). The separate checkpoint diagnostic in Appendix A reports a close but not identical Spambase comparison (0.9050 to 0.9113), so it is used only as supporting diagnostic evidence. Second, the Q1 boundary analysis shows accuracy gains after adding the Complementary module in low-confidence regions.

Table 12. Boundary-bin (Q1) complementary correction analysis. Q1 denotes the lowest-confidence quartile obtained via equal-frequency binning ($n_{\text{bins}} = 4$) of the Primary confidence within each fold. For binary tasks, $\text{conf}(\mathbf{x}) = \max(\mathbf{p}_{\text{primary}}, 1 - \mathbf{p}_{\text{primary}})$; for multi-class tasks, $\text{conf}(\mathbf{x}) = \max_c \mathbf{p}_{\text{primary}}(c)$.

Dataset	Q1 N	Acc _{Primary,Q1}	Acc _{Full,Q1}	ΔAcc (pp)	$w \rightarrow r$	$r \rightarrow w$	Error Recovery Rate
BCWD	174	0.9023	0.9138	+1.15	3	1	17.6%
Vowel	245	0.8980	0.9020	+0.41	3	2	12.0%
Spambase	1150	0.7852	0.8026	+1.74	41	21	16.6%
Gisette	1500	0.9087	0.9100	+0.13	4	2	2.9%

5.4. Additional Validation and Claim Boundaries

This section bounds the empirical claim with a compact validation block. Appendix A gives the full evidence tables and visualization diagnostics. These analyses ask four focused questions: whether the gains remain competitive against Random Forest, XGBoost, LightGBM, CatBoost, and EBM; whether the gains are specific to role-separated routing; whether sequential training is only a heuristic; and whether threshold or sparsity choices materially affect the results. The modern-baseline rows in Tables 7–10 report Accuracy,

Precision, Recall, and F1. The Appendix A also compares SHAP, aggregated LIME, and the GRS-ANFIS fuzzy rule path on the Breast Cancer dataset. These analyses bound the claim rather than infer universal dominance.

Table 13 summarizes the compact architecture attribution test. The primary GRS-ANFIS (Full) F1 values remain those in the dataset-specific performance tables: 0.9736 for Vowel, 0.9111 for Spambase, 0.9553 for Breast Cancer, and 0.9758 for Gisette (Tables 7–10). Table 13 is a separate supplemental diagnostic experiment. It compares only the compact variant set shown there: diagnostic Full GRS, no-disjoint routing, dense TSK with the same rule budget, and product firing. The results show substantial variant sensitivity. No-disjoint routing, dense same-rule TSK, or product firing can match or exceed diagnostic Full GRS-ANFIS on some datasets, and the diagnostic Full configuration can collapse under stricter hard-mask settings. Joint training and no-hard-refinement are reported separately in Appendix A Table A4. Thus, the architectural claim is not that disjoint routing always maximizes F1. The claim is that it produces a traceable primary-plus-complementary rule structure under a bounded and explicitly reported performance trade-off.

Table 13. Compact architecture ablation. Values are F1 scores from supplemental diagnostic experiments, not replacements for the primary cross-validation results in Tables 7–10. Variants should be compared within this diagnostic protocol. No-disjoint variants allow feature overlap between modules; dense TSK uses the same total rule budget without role separation; product firing replaces HTSK-style firing.

Dataset	Diagnostic Full GRS	No Disjoint Routing	Dense TSK Same Rules	Product Firing
Breast Cancer	0.0000	0.9378	0.9234	0.9578
Vowel	0.9666	0.9716	0.9635	0.9459
Spambase	0.8701	0.9125	0.8893	0.8933
Gisette	0.6667	0.9757	0.9498	0.6667

Table 14 addresses training-schedule and robustness questions. Full inference improves over the Primary checkpoint by a small but positive F1 delta in all four datasets, with zero hard-route overlap after binarization. The separate schedule diagnostic is mixed: Vowel favors the sequential diagnostic setting, whereas joint or non-refined variants are stronger in several other diagnostic runs. Hence, we describe sequential freezing as a constrained training heuristic rather than a convergence guarantee. Threshold and sparsity rows further show that performance is generally stable near $\tau = 0.5$ but can degrade or collapse at stricter thresholds, especially on Breast Cancer, Spambase, and Gisette.

Table 14. Compact schedule, threshold, and sparsity evidence. The threshold range reports the minimum–maximum F1 across $\tau \in \{0.3, 0.4, 0.5, 0.6, 0.7\}$. The $\lambda = 0.01$ column reports F1 and hard-route overlap.

Dataset	Full–Primary F1	Sequential/Joint F1	$\tau = 0.5$ F1 (Range)	$\lambda = 0.01$ F1/Overlap
Breast Cancer	+0.0043	0.0000/0.9575	0.9553 (0.0000–0.9553)	0.9508 $\pm$ 0.0049/0.0000
Vowel	+0.0011	0.9666/0.3364	0.9736 (0.9666–0.9736)	0.9684 $\pm$ 0.0116/0.0000
Spambase	+0.0063	0.8701/0.8949	0.9113 (0.8701–0.9113)	0.9125 $\pm$ 0.0153/0.0000
Gisette	+0.0003	0.6667/0.9724	0.9758 (0.6667–0.9758)	0.9752 $\pm$ 0.0056/0.0000

5.5. Case-Level Explanation of the Correction Mechanism

To show the explanation form of GRS-ANFIS more clearly, we analyze one representative Breast Cancer test sample (fold 4, case 488), whose ground-truth label is *malignant*. This sample is useful because it contains both benign-like and malignant-like evidence. As a result, it makes the internal decision process easy to inspect. The raw source-variable values are listed in Table 15.

Table 15. Raw source-variable values of the representative Breast Cancer test case (fold 4, case 488).

Feature	Value
Clump_thickness	3
Uniformity_of_cell_size	6
Uniformity_of_cell_shape	4
Marginal_adhesion	10
Single_epithelial_cell_size	3
Bare_nuclei	3
Bland_chromatin	3
Normal_nucleoli	4
Mitoses	1
Ground truth	malignant

Tables 15–17 provide a representative case-level illustration of the correction mechanism. The main Breast Cancer experiments use the 80-dimensional preprocessed model-input representation described in Section 4.1. For interpretive display, this representative sample is projected back to the nine raw diagnostic source variables, so that the Primary and Complementary correction pathway can be read at the source-variable level. Therefore, the raw-variable counts shown in this case-level illustration should not be interpreted as the model-input dimensionality or selected-feature count used in the main 5-fold experiments.

Table 16. Representative case-level prediction illustration for Breast Cancer case 488. The displayed feature count refers to raw/source variables shown for interpretability, not to the 80-dimensional preprocessed input representation used in the main experiments.

Model	Raw Variables Shown	Predicted Class	Malignant Probability	Correct
ANFIS	9	malignant	0.8788	True
GA-ANFIS	5	benign	0.0633	False
PSO-ANFIS	7	malignant	0.8488	True
H-ANFIS	9	benign	0.3117	False
GRS-ANFIS (Full)	9	malignant	0.9459	True

This case is not presented to claim that GRS-ANFIS is the only model that predicts correctly. Table 16 shows that ANFIS, PSO-ANFIS, and GRS-ANFIS all predict the malignant class correctly, whereas GA-ANFIS and H-ANFIS predict the benign class incorrectly. Therefore, the main point of this case study is different: GRS-ANFIS does not only produce the correct class, but also makes the correction mechanism visible.

Table 16 directly shows this pattern. The final output of GRS-ANFIS is a correct malignant prediction with probability 0.9459. In the internal sample-level decomposition, however, the Primary module alone is benign-oriented (Primary sigmoid score = 0.1822), while the Complementary module provides a strong malignant correction signal (sigmoid-transformed complementary score = 0.9874). Because the two branches are combined additively in score/logit space, Example 1 below converts these sample-level sigmoid scores to logits before summing them. This means that the final decision is not obtained by a single undivided rule response. Instead, the model first forms a benign-oriented interpretation and then corrects it through the Complementary module.

Table 17. Representative case-level module decomposition and IF–THEN explanation for Breast Cancer case 488.

Model/Component	Sample- or Module-Level Output	Representative IF–THEN Pattern	Interpretive Reading
ANFIS	Sample output: malignant probability = 0.8788; prediction = malignant.	IF Clump_thickness is low AND Uniformity_of_cell_size is high AND Uniformity_of_cell_shape is high AND Marginal_adhesion is high AND Single_epithelial_cell_size is low AND Bare_nuclei is low AND Bland_chromatin is low AND Normal_nucleoli is low AND Mitoses is mid(mf1) THEN malignant-oriented pattern.	The sample is correctly classified, but benign-like and malignant-like evidence is mixed inside one full-input rule. The rule does not explicitly show a correction process.
GA-ANFIS	Sample output: malignant probability = 0.0633; prediction = benign.	IF Clump_thickness is mid(mf3) AND Single_epithelial_cell_size is mid(mf2) AND Bare_nuclei is high AND Bland_chromatin is low AND Mitoses is high THEN benign-oriented pattern.	The selected 5-feature rule misses the key malignant pattern in this sample, especially the combination centered on Uniformity_of_cell_size and Marginal_adhesion. This leads to a benign misclassification.
PSO-ANFIS	Sample output: malignant probability = 0.8488; prediction = malignant.	IF Clump_thickness is mid(mf1) AND Uniformity_of_cell_size is low AND Uniformity_of_cell_shape is mid(mf2) AND Marginal_adhesion is high AND Bare_nuclei is mid(mf1) AND Bland_chromatin is mid(mf1) AND Normal_nucleoli is high THEN malignant-oriented pattern.	The sample is correctly classified through one representative rule, but the explanation still remains a single-rule result rather than an explicit correction process.
H-ANFIS	Sample output: malignant probability = 0.3117; prediction = benign.	module A: IF Marginal_adhesion is high AND Single_epithelial_cell_size is high AND Normal_nucleoli is high AND Mitoses is low; module B: IF Clump_thickness is low AND Uniformity_of_cell_size is high AND Uniformity_of_cell_shape is low AND Bare_nuclei is low AND Bland_chromatin is high.	The model separates the evidence into two modules, but the final aggregation still stays on the benign side. The decomposition is visible, but the correction fails.
GRS-ANFIS Primary	Primary module: sigmoid-transformed Primary score = 0.1822; $y_{\text{primary}} = -1.5015$.	Primary: IF Clump_thickness is low AND Uniformity_of_cell_shape is low AND Bare_nuclei is low AND Bland_chromatin is high AND Normal_nucleoli is low AND Mitoses is low.	The Primary module forms a benign-oriented initial interpretation.
GRS-ANFIS Complementary	Complementary module: sigmoid-transformed complementary score = 0.9874; $y_{\text{complementary}} = +4.3614$.	Complementary: IF Uniformity_of_cell_size is high AND Marginal_adhesion is high AND Single_epithelial_cell_size is low.	The Complementary module captures a malignant correction pattern in a distinct feature group.
GRS-ANFIS Full	Full model: $y_{\text{primary}} + y_{\text{complementary}} = +2.8599$; $\sigma(2.8599) = 0.9458$; reported full probability = 0.9459.	Primary and Complementary module outputs are combined in logit space.	The malignant correction overturns the benign-oriented Primary output and yields the final correct malignant prediction.

Table 17 reports the sample-level prediction or module-level decomposition together with representative IF–THEN patterns for the same case. For the baseline ANFIS variants, the numeric output column uses the sample-level malignant probability and prediction from Table 16, while the IF–THEN pattern provides a qualitative representative explanation. For GRS-ANFIS, the numeric output column reports the Primary, Complementary, and Full module-level decomposition used in Example 1. Thus, the table and Example 1 now use the same sample/module-level numeric layer.

For ANFIS, the representative pattern is consistent with a strong malignant sample-level prediction, but the explanation is still a single full-input rule that mixes benign-like and malignant-like cues. The rule is correct, but it does not separate the competing evidence into different roles. PSO-ANFIS shows the same limitation. Its representative pattern also leads to the correct malignant class, but the explanation remains a single-rule conclusion.

By contrast, GA-ANFIS fails because its selected rule misses the key malignant pattern in this sample. In particular, it does not capture the combination of high *Uniformity_of_cell_size* and high *Marginal_adhesion*. As a result, it produces a strongly benign-oriented output. H-ANFIS separates the evidence into two modules, which is more structured than ANFIS and PSO-ANFIS, but the final aggregation still remains on the benign side. Therefore, its decomposition is visible, but its correction is unsuccessful.

GRS-ANFIS shows the most informative explanation in Table 17. The representative Primary pattern is benign-oriented and is driven by low *Uniformity_of_cell_shape*, low *Bare_nuclei*, and low *Mitoses*. The representative Complementary pattern then focuses on a different feature group: high *Uniformity_of_cell_size*, high *Marginal_adhesion*, and low *Single_epithelial_cell_size*. The module-level values show how this Complementary branch provides a strong malignant correction and reverses the initial benign-oriented tendency. Hence, GRS-ANFIS explains not only *what* the final class is, but also *how* the final class is reached through correction.

Example 1 (empirical role-decomposition example from Breast Cancer case 488). Using the module-level values reported in Table 17, the Primary module is benign-oriented, while the Complementary module captures a malignant correction pattern driven by high *Uniformity_of_cell_size* and high *Marginal_adhesion*. The corresponding case-level decomposition is computed in logit space:

$$\begin{aligned} y_{\text{primary}} &= \text{logit}(0.1822) = -1.5015, \\ y_{\text{complementary}} &= \text{logit}(0.9874) = +4.3614, \\ \tilde{y} &= y_{\text{primary}} + y_{\text{complementary}} = +2.8599, \\ \sigma(\tilde{y}) &= \sigma(2.8599) = 0.9458. \end{aligned}$$

This reconstructed Full probability matches the reported GRS-ANFIS Full probability 0.9459 up to rounding. Thus, the case provides an empirical role-decomposition example: a benign-oriented Primary interpretation is overturned by a malignant Complementary correction using a distinct rule pathway.

Overall, this case makes the main interpretability advantage of GRS-ANFIS explicit. ANFIS and PSO-ANFIS are also correct on this sample, but they present the decision mainly as a single-rule outcome. GRS-ANFIS, in contrast, presents the decision as a two-step process: an initial benign-oriented interpretation followed by a targeted malignant correction. This makes the internal reasoning more transparent and more useful for human inspection.

6. Discussion

6.1. Summary of Empirical Findings

Across the four benchmarks, GRS-ANFIS (Full) shows strong predictive performance while preserving a role-separated fuzzy structure. Within the ANFIS-family rows, it provides the strongest F1 in all four main performance tables (Tables 7–10). The modern-baseline rows report full predictive metrics for Random Forest, XGBoost, LightGBM, CatBoost, and EBM. These non-fuzzy baselines exceed GRS-ANFIS on some predictive metrics, including SVM on Breast Cancer and LightGBM on Spambase and Gisette. We

therefore frame the result as competitive neuro-fuzzy performance with role-separated explanation, not universal predictive dominance.

Within the Q1 region (lowest-confidence quartile), adding the Complementary module produces dataset-dependent gains over Primary-only inference: +1.15 pp (BCWD), +0.41 pp (Vowel), +1.74 pp (Spambase), and +0.13 pp (Gisette). These deltas suggest that the correction stage contributes most for low-confidence samples, with the clearest boundary-bin effect on Spambase.

The additional validation tables refine this interpretation. The compact architecture ablation in Table 13 is a supplemental diagnostic experiment and should be read separately from the primary performance tables. It shows that disjoint routing is not always the F1-maximizing variant; dense same-rule TSK, non-disjoint routing, or product firing can be stronger in some compact diagnostic cases. Joint training and no-hard-refinement are reported in Appendix A Table A4 and are likewise treated as sensitivity evidence. Thus, the value of the proposed design is its explicit separation of Primary and Complementary rule roles under competitive accuracy, rather than an unconditional performance advantage. Table 14 similarly shows that schedule and threshold effects are mixed, so these settings should not be read as convergence guarantees.

6.2. Role-Separated Inference and Structured Correction

A key empirical pattern is that the Complementary module is most useful in difficult regions, not uniformly across all samples. In the Q1 analysis (Table 12), Full inference improves boundary-near accuracy by +1.15 pp on BCWD, +0.41 pp on Vowel, +1.74 pp on Spambase, and +0.13 pp on Gisette. Spambase shows the largest net correction, with 41 wrong $\rightarrow$ right flips versus 21 right $\rightarrow$ wrong flips.

Under this framing, GRS-ANFIS behaves as a structured two-stage inference system: the Primary module establishes the main decision context, and the Complementary module selectively revises uncertain cases. This is consistent with the design goal of separating coarse decision formation from conditional refinement, rather than relying on a single undifferentiated rule pool.

6.3. Interpretability and Structural Compactness

Beyond predictive accuracy, GRS-ANFIS is designed to preserve interpretability through explicit functional decomposition. The Primary models use 7 rules, and the Full models use 18 rules, which remain below the 30-rule budget used by a standard TSK-ANFIS baseline in our experiments (Tables 7–10). This indicates that improved performance does not require a proportional growth in a monolithic rule base.

For quantitative interpretability, we report a unified Nauck-based interpretability index. Specifically, ANFIS, GA-ANFIS, and PSO-ANFIS are evaluated by the Nauck index directly, whereas H-ANFIS and GRS-ANFIS are evaluated by HFSi [29] with subsystem-level Nauck indices. This allows the comparison to retain a common rule-based interpretability basis while additionally reflecting hierarchical or modular structure when applicable.

The Appendix A also compares the explanation forms of SHAP, LIME, and GRS-ANFIS. SHAP and LIME are presented as post-hoc feature-attribution methods, whereas GRS-ANFIS is presented as a model-native fuzzy rule-path explanation that exposes Primary and Complementary IF-THEN structures. Because post-hoc attribution can be vulnerable to adversarial manipulation or counterfactual fragility in some settings [30,31], these analyses are intended to clarify the different types of interpretability provided by each method, not to rank them using a single perturbation-based faithfulness criterion.

6.4. Position of HTSK in This Study

It is important to clarify that HTSK is not treated here as a competing architecture against GRS-ANFIS. Instead, HTSK-style firing computation is used as a shared training condition for all ANFIS-family models, following prior analysis on high-dimensional trainability [6]. This setup reduces confounding from numerical firing instability and makes the comparison focus on inference organization (standard TSK-ANFIS vs. role-separated), not on whether one model happened to use a more stable firing formula.

Accordingly, the improvements observed for GRS-ANFIS should be interpreted as evidence for role-separated modeling under a common stabilized fuzzy-learning regime. The product-vs.-HTSK checks in Appendix A support this narrower reading. HTSK can help avoid unstable firing behavior in high-dimensional settings, but it is not presented as the sole source of the observed performance. The product-firing values in the compact architecture ablation and the focused HTSK-vs.-product diagnostic come from different targeted checks. They therefore support separate claims and should not be treated as a single shared estimate.

6.5. Limitations and Future Work

This study has several limitations.

- Correction-module complexity: The Complementary module improves difficult cases but increases rule count relative to Primary-only models; tighter sparsity/pruning constraints are needed.
- High-dimensional interpretability: In extreme-dimensional settings (e.g., Gisette), interpretability indices can become less discriminative, motivating additional complementary measures.
- Boundary-analysis sensitivity: Q1 results depend on confidence binning and calibration choices; alternative partitions and calibrated uncertainty estimates should be studied.
- Ablation scope: The validation experiments include architecture, schedule, threshold, sparsity, product-vs.-HTSK, correlated-feature, and explanation-form diagnostics, but they are still targeted analyses rather than an exhaustive design-space search.
- Supplemental diagnostic comparability: Some validation tables use targeted supplemental diagnostic experiments rather than the primary cross-validation result tables. Therefore, variant comparisons should be made within the same diagnostic table unless the text explicitly states otherwise.
- Robustness and stability: Stage-wise training with hard-gate freezing may depend on thresholds and random seeds. The supplemental sensitivity tables reduce this concern but do not provide theoretical guarantees.
- Modern baseline boundary: Default-hyperparameter Random Forest and XGBoost, and supplemental baselines such as LightGBM, CatBoost, and EBM, exceed GRS-ANFIS on some predictive metrics, while they do not provide Nauck-based rule interpretability or intrinsic role-separated explanations. The claim is therefore bounded to competitive neuro-fuzzy modeling with intrinsically role-separated explanations.

Future work will focus on adaptive gate budgeting, module-specific regularization, and larger multi-domain benchmarks to better quantify when role-separated fuzzy correction is most beneficial.

7. Conclusions

This study presented GRS-ANFIS, a role-separated neuro-fuzzy framework that combines consequent-only soft gating with hard subset routing. Under a shared HTSK-style learning condition, the four benchmark experiments indicate that GRS-ANFIS can

maintain competitive predictive performance with compact, structured rule bases. The Complementary-module gains are measurable but dataset-dependent in low-confidence regions. This supports the intended interpretation of targeted correction rather than global replacement of the Primary inference. The fixed module-specific routing masks also provide explicit variable-role assignments that can be inspected directly.

The main contribution of GRS-ANFIS is an interpretability-oriented architecture. It separates inference into Primary decision formation and Complementary correction, converts learned gates into inspectable hard routing masks, and provides a compact neuro-fuzzy rule representation. The proposed model should therefore be viewed as a role-separated fuzzy architecture with competitive performance and explicit structural explanations, not as a state-of-the-art predictive classifier for all tabular settings.

Several methodological limitations remain. Interpretability gains are dataset-dependent, especially in extremely high-dimensional settings where rule-based interpretability indices may be less informative. The estimated correction effect also depends on confidence partitioning, thresholds, and initialization. The additional validation analyses show that GRS-ANFIS is competitive rather than universally dominant. They also show that sequential freezing and hard-threshold choices are sensitivity-dependent heuristics, not convergence guarantees. Future work should include broader ablations, richer case-level explanation examples, and tighter complexity control for correction rules.

Author Contributions: Conceptualization, J.H.L. and S.J.; methodology, J.H.L. and S.J.; software, J.H.L.; validation, J.H.L., S.K., and S.J.; formal analysis, J.H.L. and S.K.; investigation, J.H.L.; resources, S.J.; data curation, J.H.L.; writing—original draft preparation, J.H.L.; writing—review and editing, S.K. and S.J.; visualization, J.H.L.; supervision, S.J.; project administration, S.J.; funding acquisition, S.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partly supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP)-ITRC (Information Technology Research Center) grant funded by the Korea government (MSIT) (IITP-2026-RS-2020-II201808, 50%) and by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: RS-2022-KH130593, 50%). The funders had no role in the design of the study, the collection, analysis, and interpretation of the data, or the writing of the manuscript.

Informed Consent Statement: Not applicable.

Data Availability Statement: All experiments were conducted on publicly available benchmark datasets. No new dataset was created for this study. Breast Cancer Wisconsin, Spambase, Vowel, and Gisette are available from their respective public benchmark sources. The implementation code, configuration files, and reproduction notebooks without execution outputs are publicly available at <https://github.com/harp3133t/GRS-ANFIS> (accessed on 3 April 2026).

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Supplementary Validation Evidence

This appendix records the additional validation evidence used to support bounded claims in the main text. GRS-ANFIS is competitive as a role-separated neuro-fuzzy model, but the additional evidence does not support a claim of universal dominance over all tabular baselines or all architectural variants.

Table A1. Threshold sensitivity across $\tau \in \{0.3, 0.4, 0.5, 0.6, 0.7\}$. Selected counts are encoded/preprocessed input dimensions used by gate masks; Breast Cancer uses 80 one-hot/preprocessed input dimensions from 9 raw variables, and Vowel uses 29 one-hot/preprocessed input dimensions from 27 raw variables in these supplemental analyses.

Dataset	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$	Selected Dims at 0.5	Mask Jaccard at 0.5
Breast Cancer	0.9532	0.9553	0.9553	0.9096	0.0000	76.2	1.0000
Vowel	0.9736	0.9736	0.9736	0.9707	0.9666	29.0	1.0000
Spambase	0.9070	0.9096	0.9113	0.8855	0.8701	53.4	1.0000
Gisette	0.9757	0.9757	0.9758	0.6667	0.6667	2680.8	1.0000

Table A2. Sparsity coefficient and seed sensitivity summary. Values report mean $\pm$ standard deviation where available. The displayed scalar λ denotes the diagnostic setting $\lambda_{\text{primary}, S1} = \lambda_{\text{complementary}, S2}$; it is not a replacement for the dataset-specific Primary/Complementary settings in Table 6. Selected counts are encoded/preprocessed input dimensions; Breast Cancer counts are out of 80 one-hot/preprocessed dimensions and Vowel counts are out of 29 one-hot/preprocessed dimensions in these supplemental analyses.

Dataset	λ	Full Acc	Full F1	Selected Dims	Mask Jaccard	P/C Overlap
Breast Cancer	0	0.9628 $\pm$ 0.0032	0.9470 $\pm$ 0.0040	77.6000 $\pm$ 1.1402	0.9418	0.0000
Breast Cancer	0.01	0.9657 $\pm$ 0.0032	0.9508 $\pm$ 0.0049	70.4000 $\pm$ 1.5166	0.9859	0.0000
Breast Cancer	0.1	0.9499 $\pm$ 0.0142	0.9262 $\pm$ 0.0216	51.2000 $\pm$ 1.3038	0.7509	0.0000
Vowel	0	0.9747 $\pm$ 0.0118	0.9747 $\pm$ 0.0117	29.0000 $\pm$ 0.0000	1.0000	0.0000
Vowel	0.01	0.9687 $\pm$ 0.0115	0.9684 $\pm$ 0.0116	28.6000 $\pm$ 0.5477	0.9862	0.0000
Vowel	0.1	0.9566 $\pm$ 0.0150	0.9565 $\pm$ 0.0149	21.2000 $\pm$ 1.4832	0.7310	0.0000
Spambase	0	0.9309 $\pm$ 0.0085	0.9097 $\pm$ 0.0123	54.8000 $\pm$ 1.3038	0.7978	0.0000
Spambase	0.01	0.9326 $\pm$ 0.0110	0.9125 $\pm$ 0.0153	40.6000 $\pm$ 1.1402	0.6883	0.0000
Spambase	0.1	0.9172 $\pm$ 0.0116	0.8903 $\pm$ 0.0167	13.8000 $\pm$ 0.8367	0.2345	0.0000
Gisette	0	0.9762 $\pm$ 0.0034	0.9761 $\pm$ 0.0034	4532.0000 $\pm$ 14.8661	0.5753	0.0000
Gisette	0.01	0.9752 $\pm$ 0.0057	0.9752 $\pm$ 0.0056	2416.8000 $\pm$ 20.2163	0.9766	0.0000
Gisette	0.1	0.9713 $\pm$ 0.0067	0.9713 $\pm$ 0.0068	362.6000 $\pm$ 12.5817	0.1505	0.0000

Table A3. Full-vs.-Primary checkpoint attribution and sequential-vs.-joint schedule comparison. Values are from the checkpoint diagnostic protocol and should be read separately from the primary dataset-specific performance tables; small differences from Tables 7–10 reflect this diagnostic scope.

Dataset	Primary F1	Full F1	Delta	Hard Overlap	Sequential F1	Joint No-Freeze F1
Breast Cancer	0.9510	0.9553	+0.0043	0.0	0.0000	0.9575
Vowel	0.9725	0.9736	+0.0011	0.0	0.9666	0.3364
Spambase	0.9050	0.9113	+0.0063	0.0	0.8701	0.8949
Gisette	0.9755	0.9758	+0.0003	0.0	0.6667	0.9724

Table A4. Architecture ablation summary. Values are F1 scores from supplemental diagnostic experiments and are intended for within-table variant comparison; they do not replace the primary cross-validation performance values reported in Tables 7–10. Parentheses in the no-disjoint column report overlap rates.

Dataset	Diagnostic Full GRS	Joint	No Disjoint	No Hard Fine-Tune	Dense Same Rules	Product Firing
Breast Cancer	0.0000	0.9575	0.9378 (0.3638)	0.9594	0.9234	0.9578
Vowel	0.9666	0.3364	0.9716 (0.3696)	0.9697	0.9635	0.9459
Spambase	0.8701	0.8949	0.9125 (0.0683)	0.8998	0.8893	0.8933
Gisette	0.6667	0.9724	0.9757 (0.2578)	0.9696	0.9498	0.6667

Table A5. Focused HTSK-style firing versus product firing diagnostic. Values come from a separate firing-mode check and should not be conflated with the architecture-ablation diagnostics in Table A4. HTSK is treated as a numerical scale-control setting, not as the sole cause of GRS-ANFIS performance.

Dataset	HTSK F1	Product F1	HTSK Base Near-Zero	Product Base Near-Zero	HTSK Comp Rules	Product Comp Rules
Breast Cancer	0.9553 ± 0.0108	0.9578 ± 0.0041	0.0000 ± 0.0000	0.0053 ± 0.0042	10.9795 ± 0.0050	10.9401 ± 0.1340
Vowel	0.9736 ± 0.0057	0.9459 ± 0.0121	0.0000 ± 0.0000	0.0000 ± 0.0000	11.0000 ± 0.0000	11.0000 ± 0.0000
Spambase	0.9113 ± 0.0059	0.8933 ± 0.0072	0.0000 ± 0.0000	0.0879 ± 0.1965	10.4088 ± 0.0905	4.2178 ± 0.1754
Gisette	0.9758 ± 0.0047	0.6667 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	11.0000 ± 0.0000	11.0000 ± 0.0000

Table A6. Correlated-feature stress test. Five copied features are appended and the table reports F1 and selected-dimension changes under the redundant-feature condition. The result supports group-level role sensitivity checks under redundant copies, not unique feature identity under collinearity.

Dataset	Copied Features	Base F1	Augmented F1	F1 Delta	Base Selected	Augmented Selected
Breast Cancer	5	0.9553 ± 0.0108	0.9467 ± 0.0082	−0.0085	76.2000 ± 0.8367	80.0000 ± 1.4142
Vowel	5	0.9736 ± 0.0057	0.9707 ± 0.0103	−0.0029	29.0000 ± 0.0000	34.0000 ± 0.0000
Spambase	5	0.9113 ± 0.0059	0.9094 ± 0.0109	−0.0020	53.4000 ± 0.5477	59.2000 ± 0.8367
Gisette	5	0.9758 ± 0.0047	0.9762 ± 0.0067	+0.0004	2680.8000 ± 33.9293	2700.4000 ± 37.0783

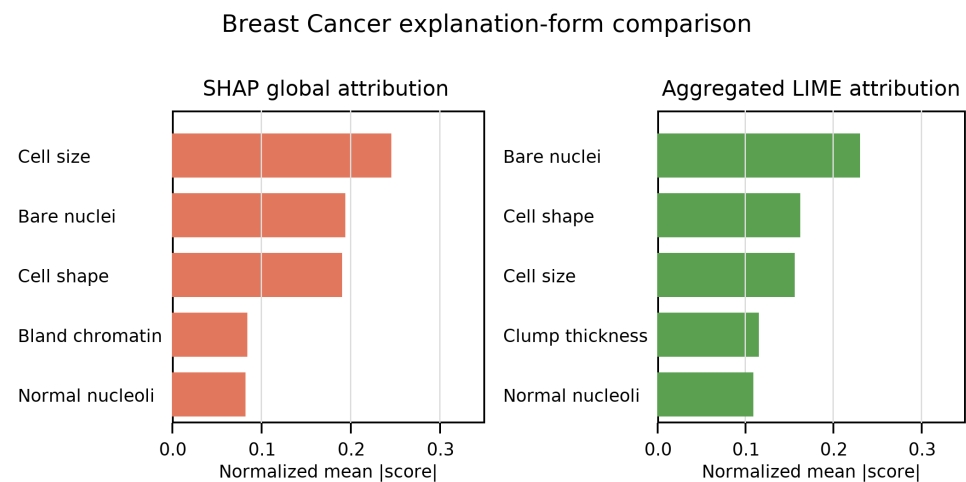
Table A7. Comparison of explanation forms between post-hoc attribution methods and GRS-ANFIS rule-path explanations.

Method	Type	Output	Interaction	Rule Path
SHAP	Post-hoc attribution	Feature contribution scores	Limited in standard summary	No
LIME	Local surrogate attribution	Local linear weights	Limited	No
GRS-ANFIS	Intrinsic rule path	Primary/Complementary IF-THEN rules	Yes, via antecedent conjunctions	Yes

Post-Hoc Attribution and Fuzzy Rule-Path Visualization

To provide a visual comparison between post-hoc explanations and the GRS-ANFIS rule interpretation, Figure A1 summarizes the global explanation forms obtained on the Breast Cancer dataset. SHAP is summarized by mean absolute feature attribution, LIME is summarized by aggregated local feature weights across validation samples, and GRS-ANFIS is summarized by its Primary/Complementary hard masks and representative IF-THEN rule paths. The figure uses the same module-level case-488 output labels as Table 17: the representative path is qualitative, while the reported Primary, Complementary, and Full outputs are sample/module-level values. The goal of this comparison is to clarify what each explanation family provides: SHAP and aggregated LIME identify influential variables, whereas GRS-ANFIS exposes how selected variables are organized into Primary and Complementary fuzzy rules.

Figure A1 provides a qualitative and global explanation-form comparison on the Breast Cancer dataset. The SHAP panel reports mean absolute SHAP values from a Random Forest classifier trained on the same Breast Cancer data, and the LIME panel reports mean absolute local weights aggregated across validation samples. These post-hoc panels show which variables are influential in a standard feature-attribution format. The GRS-ANFIS panel instead shows the hard role masks and the representative Primary/Complementary fuzzy-rule path, emphasizing how the selected variables are combined into an explicit correction process. The numeric labels in the GRS-ANFIS panel follow the module-level decomposition used in Table 17 and Example 1. This comparison supports a complementary interpretation of SHAP/LIME and GRS-ANFIS rather than a universal superiority claim for either explanation family.



GRS-ANFIS module-level structure and representative case-488 path

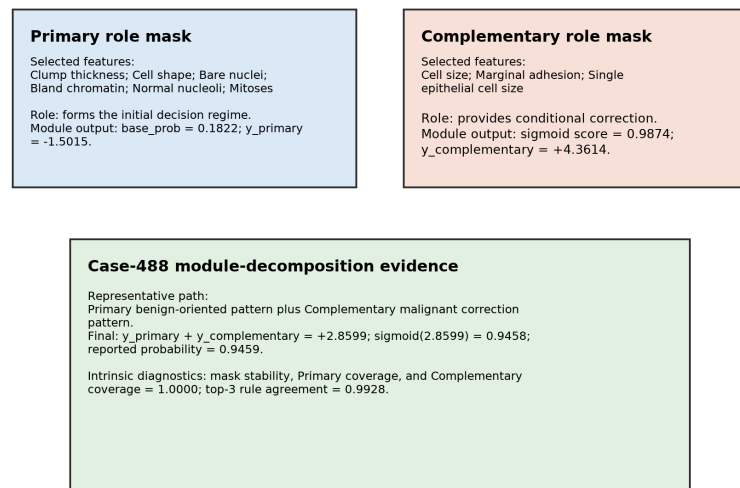


Figure A1. Global explanation-form comparison on the Breast Cancer dataset. SHAP and aggregated LIME provide feature-attribution summaries, whereas GRS-ANFIS provides a model-native fuzzy rule path with separate Primary and Complementary roles. The GRS-ANFIS numeric labels use the same case-488 module-level decomposition as Table 17.

Figure A2 visualizes the explanation form for the representative Breast Cancer case analyzed in Section 5.5. The left panel shows the post-hoc explanation format as feature attribution, while the right panel shows the GRS-ANFIS explanation format as a Primary rule followed by a Complementary correction rule. This figure is a schematic explanation-form visualization tied to the representative IF–THEN patterns and module-level decomposition in Table 17, not a new numeric estimate of per-feature SHAP or LIME values. Its Primary, Complementary, and Full values are the same sample/module-level values reported in Table 17 and Example 1.

Case 488 explanation forms: module-level decomposition

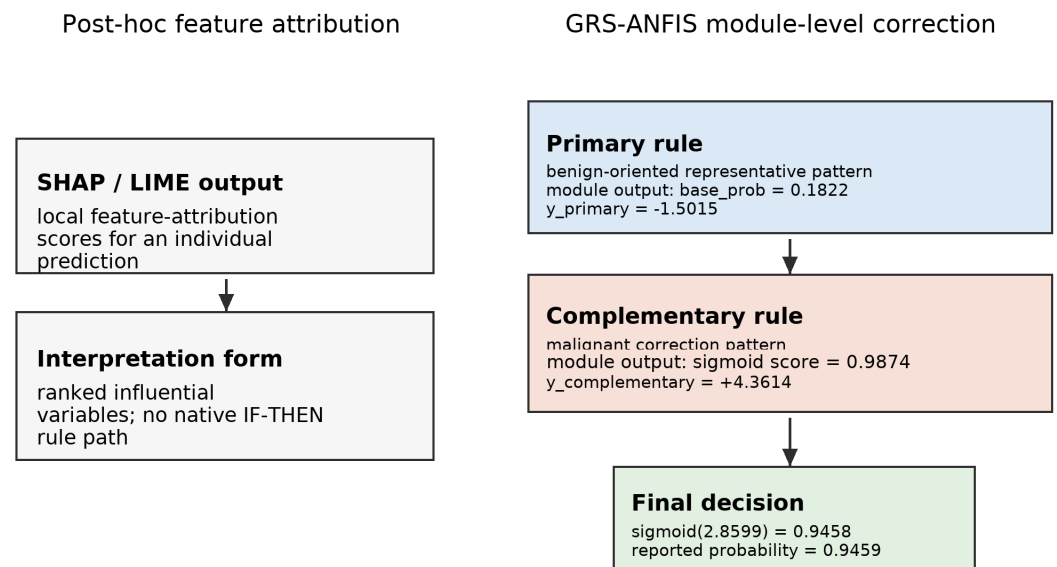


Figure A2. Explanation-form comparison for Breast Cancer case 488. Post-hoc attribution provides feature-importance style explanations; GRS-ANFIS exposes a fuzzy rule path in which a benign-oriented Primary rule is corrected by a malignant Complementary rule. The displayed outputs are module-level values: Primary sigmoid score = 0.1822, sigmoid-transformed complementary score = 0.9874, and final probability 0.9459.

Table A8. Supplemental tabular-baseline full-metric summary for the LightGBM, CatBoost, and EBM comparators. Values are mean $\pm$ std across the 5-fold supplemental baseline analyses.

Dataset	Model	Acc	Precision	Recall	F1
Breast Cancer	LightGBM	0.9571 $\pm$ 0.0113	0.9454 $\pm$ 0.0177	0.9296 $\pm$ 0.0310	0.9371 $\pm$ 0.0168
Breast Cancer	CatBoost	0.9599 $\pm$ 0.0108	0.9499 $\pm$ 0.0230	0.9335 $\pm$ 0.0274	0.9413 $\pm$ 0.0163
Breast Cancer	EBM	0.9671 $\pm$ 0.0082	0.9519 $\pm$ 0.0356	0.9544 $\pm$ 0.0226	0.9525 $\pm$ 0.0113
Vowel	LightGBM	0.9475 $\pm$ 0.0216	0.9499 $\pm$ 0.0219	0.9475 $\pm$ 0.0216	0.9472 $\pm$ 0.0216
Vowel	CatBoost	0.9071 $\pm$ 0.0181	0.9097 $\pm$ 0.0187	0.9071 $\pm$ 0.0181	0.9063 $\pm$ 0.0183
Vowel	EBM	0.9091 $\pm$ 0.0211	0.9117 $\pm$ 0.0209	0.9091 $\pm$ 0.0211	0.9078 $\pm$ 0.0218
Spambase	LightGBM	0.9550 $\pm$ 0.0066	0.9428 $\pm$ 0.0088	0.9432 $\pm$ 0.0169	0.9429 $\pm$ 0.0087
Spambase	CatBoost	0.9511 $\pm$ 0.0057	0.9467 $\pm$ 0.0071	0.9283 $\pm$ 0.0166	0.9373 $\pm$ 0.0079
Spambase	EBM	0.9509 $\pm$ 0.0068	0.9426 $\pm$ 0.0064	0.9321 $\pm$ 0.0202	0.9372 $\pm$ 0.0094
Gisette	LightGBM	0.9790 $\pm$ 0.0037	0.9793 $\pm$ 0.0061	0.9787 $\pm$ 0.0018	0.9790 $\pm$ 0.0037
Gisette	CatBoost	0.9725 $\pm$ 0.0061	0.9721 $\pm$ 0.0095	0.9730 $\pm$ 0.0043	0.9725 $\pm$ 0.0060
Gisette	EBM	0.9738 $\pm$ 0.0050	0.9737 $\pm$ 0.0084	0.9740 $\pm$ 0.0056	0.9738 $\pm$ 0.0049

Table A9. Paired statistics for selected comparisons in the supplemental validation evidence. Deltas are GRS-ANFIS (Full) minus the comparator. The reported p -values are two-sided paired t -tests computed on the five fold-level metric differences and are used as exploratory evidence because only five paired folds are available.

Dataset	Model	Acc	Precision	Recall	F1
Breast Cancer	ANFIS	+0.0215	<0.001	+0.0318	<0.001
Breast Cancer	Random Forest	+0.0072	0.0340	+0.0115	0.0193
Breast Cancer	XGBoost	+0.0143	0.1422	+0.0229	0.1257
Vowel	GA-ANFIS	+0.2061	<0.001	+0.2094	<0.001
Vowel	PSO-ANFIS	+0.1444	<0.001	+0.1461	<0.001
Vowel	H-ANFIS	+0.0424	<0.001	+0.0426	<0.001
Vowel	SVM	+0.1172	<0.001	+0.1192	<0.001
Vowel	Random Forest	+0.0020	0.5870	+0.0021	0.5766
Vowel	XGBoost	+0.0606	0.0100	+0.0611	0.0095

Table A9. Cont.

Dataset	Model	Acc	Precision	Recall	F1
Spambase	GRS-ANFIS(base)	+0.0043	0.0993	+0.0064	0.0731
Spambase	GA-ANFIS	+0.0220	0.0012	+0.0324	0.0012
Spambase	PSO-ANFIS	+0.0461	<0.001	+0.0690	<0.001
Spambase	H-ANFIS	+0.0180	<0.001	+0.0257	<0.001
Spambase	SVM	+0.2197	<0.001	+0.3599	<0.001
Spambase	Random Forest	-0.0204	<0.001	-0.0271	<0.001
Spambase	XGBoost	-0.0193	0.0051	-0.0260	0.0054
Gisette	ANFIS	+0.0102	0.0011	+0.0104	0.0011
Gisette	GA-ANFIS	+0.0462	<0.001	+0.0460	<0.001
Gisette	PSO-ANFIS	+0.0502	<0.001	+0.0501	<0.001
Gisette	H-ANFIS	+0.0045	0.0403	+0.0046	0.0363
Gisette	Random Forest	+0.0037	0.0269	+0.0040	0.0221
Gisette	XGBoost	+0.0002	0.9181	+0.0002	0.9080

References

- Rudin, C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nat. Mach. Intell.* **2019**, *1*, 206–215. [\[CrossRef\]](#) [\[PubMed\]](#)
- Doshi-Velez, F.; Kim, B. Towards a Rigorous Science of Interpretable Machine Learning. *arXiv* **2017**, arXiv:1702.08608.
- Ribeiro, M.T.; Singh, S.; Guestrin, C. Why Should I Trust You?: Explaining the Predictions of Any Classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), San Francisco, CA, USA, 13–17 August 2016; pp. 1135–1144. [\[CrossRef\]](#)
- Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4765–4774.
- Jang, J.-S.R. ANFIS: Adaptive-Network-Based Fuzzy Inference System. *IEEE Trans. Syst. Man Cybern.* **1993**, *23*, 665–685. [\[CrossRef\]](#)
- Cui, Y.; Wu, D.; Xu, Y. Curse of Dimensionality for TSK Fuzzy Neural Networks: Explanation and Solutions. *arXiv* **2021**, arXiv:2102.04271.
- Lee, M.-L.; Chung, H.-Y.; Yu, F.-M. Modeling of Hierarchical Fuzzy Systems. *Fuzzy Sets Syst.* **2003**, *138*, 343–361. [\[CrossRef\]](#)
- Razak, T.R.; Fauzi, S.S.M.; Gining, R.A.J.M.; Ismail, M.H.; Maskat, R. Hierarchical Fuzzy Systems: Interpretability and Complexity. *Indones. J. Electr. Eng. Inform.* **2021**, *9*, 478–489. [\[CrossRef\]](#)
- Bhattacharya, M.; Das, A. Genetic Algorithm Based Feature Selection in a Recognition Scheme Using Adaptive Neuro Fuzzy Techniques. *Int. J. Comput. Commun. Control* **2010**, *5*, 458–468. [\[CrossRef\]](#)
- Elbaz, K.; Shen, S.-L.; Sun, W.-J.; Yin, Z.-Y.; Zhou, A. Prediction Model of Shield Performance During Tunneling via Incorporating Improved Particle Swarm Optimization Into ANFIS. *IEEE Access* **2020**, *8*, 39659–39671. [\[CrossRef\]](#)
- Xue, G.; Chang, Q.; Wang, J.; Zhang, K.; Pal, N.R. An Adaptive Neuro-Fuzzy System With Integrated Feature Selection and Rule Extraction for High-Dimensional Classification Problems. *IEEE Trans. Fuzzy Syst.* **2023**, *31*, 2167–2181. [\[CrossRef\]](#)
- Tyrovolas, M.; Liang, X.S.; Stylios, C. Information Flow-Based Fuzzy Cognitive Maps with Enhanced Interpretability. *Granul. Comput.* **2023**, *8*, 2021–2038. [\[CrossRef\]](#)
- Zhang, K.; Shao, T.; Sun, Y.; Xu, X.; Zhang, X.; Zhou, X.; Ding, K.; Huang, S. Interpretable Research of Fuzzy Methods: A Literature Survey. *Inf. Fusion* **2026**, *126*, 103524. [\[CrossRef\]](#)
- Varshney, A.K.; Torra, V. Literature Review of the Recent Trends and Applications in Various Fuzzy Rule-Based Systems. *Int. J. Fuzzy Syst.* **2023**, *25*, 2163–2186. [\[CrossRef\]](#)
- Jacobs, R.A.; Jordan, M.I.; Nowlan, S.J.; Hinton, G.E. Adaptive Mixtures of Local Experts. *Neural Comput.* **1991**, *3*, 79–87. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hazimeh, H.; Zhao, Z.; Chowdhery, A.; Sathiamoorthy, M.; Chen, Y.; Mazumder, R.; Hong, L.; Chi, E.H. DSelect-k: Differentiable Selection in the Mixture of Experts with Applications to Multi-Task Learning. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 29335–29347. [\[CrossRef\]](#)
- Mangasarian, O.L.; Street, W.N.; Wolberg, W.H. Breast Cancer Diagnosis and Prognosis via Linear Programming. *Oper. Res.* **1995**, *43*, 570–577. [\[CrossRef\]](#)
- Deterding, D.; Niranjan, M.; Robinson, T. *Connectionist Bench: Vowel Recognition [Dataset]*; UCI Machine Learning Repository: Irvine, CA, USA, 1988.
- Hopkins, M.; Reeber, E.; Forman, G.; Suermondt, J. *Spambase [Dataset]*; UCI Machine Learning Repository: Irvine, CA, USA, 1999. [\[CrossRef\]](#)

20. Guyon, I.; Gunn, S.; Ben-Hur, A.; Dror, G. *Gisette [Dataset]*; UCI Machine Learning Repository: Irvine, CA, USA, 2004. [[CrossRef](#)]
21. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
22. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
23. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [[CrossRef](#)]
24. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3146–3154.
25. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased Boosting with Categorical Features. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 6638–6648.
26. Lou, Y.; Caruana, R.; Gehrke, J.; Hooker, G. Accurate Intelligible Models with Pairwise Interactions. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), Chicago, IL, USA, 11–14 August 2013; pp. 623–631. [[CrossRef](#)]
27. Nori, H.; Jenkins, S.; Koch, P.; Caruana, R. InterpretML: A Unified Framework for Machine Learning Interpretability. *arXiv* **2019**, arXiv:1909.09223.
28. Nauck, D. Measuring Interpretability in Rule-Based Classification Systems. In Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), St. Louis, Missouri, USA, 25–28 May 2003; pp. 196–201.
29. Razak, T.R.; Garibaldi, J.M.; Wagner, C.; Pourabdollah, A.; Soria, D. Interpretability Indices for Hierarchical Fuzzy Systems. In Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Naples, Italy, 9–12 July 2017. [[CrossRef](#)]
30. Slack, D.; Hilgard, S.; Jia, E.; Singh, S.; Lakkaraju, H. Fooling LIME and SHAP: Adversarial Attacks on Post Hoc Explanation Methods. In Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics, and Society (AIES), Madrid, Spain, 20–22 October 2020; pp. 180–186. [[CrossRef](#)]
31. Käsbohrer, C.C.; Mair, S.; Jiang, L. Assessing the Fragility of SHAP-Based Model Explanations Using Counterfactuals. In Proceedings of the 7th Northern Lights Deep Learning Conference (NLDL), PMLR 307, Tromsø, Norway, 6–8 January 2026; pp. 211–234.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.