Journix: Journal of Informatics and Computing
Vol. 1 No. 3, 2025 : 165-180
ISSN : 3090-6784 | DOI : 10.63866/journix.v1i3.21

Integration of Fuzzy Logic and Neural Networks for Explainable Early Diagnosis of Rice Plant Diseases

Teguh Ansyor Lorosae [1*], Miftahul Jannah [1], Siti Mutmainah [1], Fathir [1], Hilyatul Mustafidah [1]

[1] Universitas Muhammadiyah Bima

Email: teguhansyor@umbima.ac.id
(* : corresponding author)

ABSTRACT − Early diagnosis of rice leaf diseases remains challenging due to subtle symptom manifestation, uncontrolled illumination, heterogeneous backgrounds, and the limited interpretability of purely data-driven models. This study proposes an explainable hybrid framework integrating a Mamdani Fuzzy Inference System (FIS) with an Artificial Neural Network (ANN) for early rice leaf disease diagnosis under real-field conditions. The framework combines engineered symptom descriptors extracted from segmented leaf regions (GLCM texture and HSV color features), acquisition-time environmental measurements, and a fuzzy-derived disease severity cue to mitigate symptom ambiguity while preserving rule-based interpretability. Experiments were conducted on 8,000 field-acquired rice leaf images collected from multiple locations, covering Healthy, bacterial leaf blight, brown spot, and leaf smut classes. Evaluation followed a leakage-controlled, location-disjoint protocol. Across five independent runs, the proposed FIS–ANN achieved an average accuracy of 91.3 ± 0.6% and a macro-F1 score of 90.8 ± 0.7%, significantly outperforming a feature-based ANN and a fine-tuned ResNet-18 baseline (paired McNemar test, $p < 0.05$). Per-class analysis shows consistent recall improvements for visually overlapping diseases, and additional evaluation on mild-severity samples confirms maintained sensitivity at early disease stages. Field deployment experiments using smartphone-acquired images from unseen locations further demonstrate robust generalization with low on-device inference latency. These results indicate that integrating fuzzy severity reasoning into a lightweight neural classifier provides a practical balance between performance, interpretability, and computational efficiency, supporting early disease screening and mobile decision-support applications in precision agriculture.

KEYWORDS: Rice Leaf Disease, Fuzzy Inference System, Neural Network, Early Diagnosis, XAI

Integrasi Logika Fuzzy dan Jaringan Syaraf Tiruan untuk Diagnosis Dini Penyakit Tanaman Padi yang Dapat Dijelaskan

ABSTRAK − Diagnosis dini penyakit daun padi masih menjadi tantangan akibat gejala awal yang samar, pencahayaan lapangan yang tidak terkendali, latar belakang yang heterogen, serta keterbatasan interpretabilitas pada model yang sepenuhnya berbasis data. Penelitian ini mengusulkan sebuah kerangka hibrida yang dapat dijelaskan dengan mengintegrasikan Mamdani Fuzzy Inference System (FIS) dan Artificial Neural Network (ANN) untuk diagnosis dini penyakit daun padi pada kondisi lapangan nyata. Kerangka ini mengombinasikan deskriptor gejala hasil rekayasa dari citra daun tersegmentasi (fitur tekstur GLCM dan warna HSV), pengukuran lingkungan saat akuisisi, serta isyarat tingkat keparahan penyakit berbasis fuzzy untuk mengurangi ambiguitas gejala sekaligus mempertahankan interpretabilitas berbasis aturan. Eksperimen dilakukan pada 8.000 citra daun padi yang dikumpulkan dari berbagai lokasi lapangan dan mencakup kelas Sehat, hawar daun bakteri, bercak cokelat, dan gosong daun. Evaluasi dilakukan menggunakan protokol leakage-controlled dengan pemisahan data berbasis lokasi. Pada lima pengujian independen, model FIS–ANN yang diusulkan mencapai akurasi rata-rata sebesar 91,3 ± 0,6% dan nilai macro-F1 sebesar 90,8 ± 0,7%, serta secara signifikan mengungguli ANN berbasis fitur dan model ResNet-18 yang telah di-*fine-tune* (p < 0,05). Analisis per kelas menunjukkan peningkatan sensitivitas yang konsisten pada penyakit dengan gejala visual yang saling tumpang tindih, dan evaluasi tambahan pada sampel tahap awal mengonfirmasi kemampuan diagnosis dini. Pengujian implementasi lapangan menggunakan citra dari smartphone di lokasi yang belum pernah dilihat sebelumnya juga menunjukkan kemampuan generalisasi yang baik dengan latensi inferensi yang rendah.

KATA KUNCI: Penyakit Daun Padi, Sistem Inferensi Fuzzy, Jaringan Saraf Tiruan, Diagnosis Dini, XAI

Received : 14-09-2025	**Revised :** 10-12-2025	**Published :** 31-12-2025

1. INTRODUCTION

Rice foliar diseases—particularly bacterial leaf blight (BLB), brown spot (BS), and leaf smut (LS)—remain major constraints on rice production in Indonesia and can lead to substantial yield losses if not detected at an early stage. In practice, disease identification still relies heavily on manual visual inspection, which is prone to subjectivity and inconsistent expertise, especially in rural areas; this limitation has motivated increasing adoption of AI-based diagnostic methods. Prior studies show that hybrid soft-computing approaches can be effective: a Fuzzy Inference System (FIS) combined with a Complex-Valued Neural Network (CVNN) achieved 92% accuracy for rice leaf disease classification and severity assessment by integrating expert fuzzy rules with neural pattern recognition [1]. Similarly, a pure fuzzy approach using image feature extraction and fuzzy reasoning reported accuracy exceeding 91%, offering interpretability through linguistic rules that are easier to communicate to practitioners [2]. On the other hand, fully data-driven models can reach very high predictive performance—for example, an optimized ANN achieved 97.94% accuracy for rice blast detection—highlighting the strength of neural learning for robust classification [3]. Collectively, these findings indicate that integrating fuzzy logic with neural networks offers a promising pathway to balance uncertainty handling and explainability with strong discriminative learning, which is particularly relevant for farmer-oriented decision support in diverse Indonesian field conditions [1], [2], [3].

Manual visual inspection remains the predominant approach for plant disease diagnosis in smallholder farming systems, yet its reliability is often compromised by limited expert availability, subjective interpretation, and uncontrolled field conditions (e.g., variable illumination, occlusions, and complex backgrounds). These constraints are especially problematic at the early stage of infection, where symptoms may manifest only as slight discoloration, small lesions, or subtle texture changes that are difficult to distinguish consistently without specialist training [4], [5]. Empirical evidence further shows that visual surveys can be time-consuming and error-prone, with performance strongly dependent on the observer and the symptom type; sensitivity and specificity may vary substantially across surveyors, undermining dependable detection and timely outbreak management [5], [6], [7]. To mitigate these limitations, automated image-based approaches leveraging machine learning and deep learning—particularly CNN-based classifiers and advanced feature extraction pipelines—have been widely explored to provide more consistent and scalable diagnosis [6], [8], [9]. Nevertheless, deploying such methods in real-world settings remains challenging due to domain shift between laboratory and field imagery, as well as the need for interpretability, accessibility, and user trust for practical adoption among smallholder farmers [10].

Recent progress in explainable artificial intelligence (XAI) has sought to improve the transparency and practical trustworthiness of deep learning–based plant disease diagnosis, addressing the frequent criticism that CNN predictions are difficult to justify to end users. Common post-hoc explanation techniques—such as LIME, Grad-CAM, and Grad-CAM++—provide visual rationales by highlighting image regions that most influence the predicted class, enabling agronomists to verify whether the model attends to biologically plausible lesion patterns rather than spurious background cues [11], [12], [13]. Beyond visualization, recent studies have explored hybrid explainable pipelines that integrate CNNs with ensemble strategies or additional generative/transformer-based components to improve diagnostic robustness while providing more explicit decision support, including narrative justifications and management recommendations [13], [14]. Several explainable deep models report high classification performance—often above 97% on curated benchmarks—suggesting that accuracy and interpretability can be jointly pursued when explanations are incorporated into the modeling workflow [11], [12], [15]. However, evidence also indicates that CNN explanations primarily reveal correlated visual cues (e.g., disease-specific color and texture patterns) and may still require careful validation for reliability and deployment readiness in real-field contexts [16], [17]. Consequently, there remains a need for diagnostic systems that combine strong predictive performance with explanations that are not only visual but also explicit and rule-consistent, facilitating actionable understanding in operational agricultural settings.

Soft computing approaches—most notably fuzzy logic—provide an interpretable reasoning framework by representing ambiguous disease symptoms through linguistic variables and expert-derived rules, thereby accommodating uncertainty and imprecision commonly encountered in agricultural diagnosis. However, purely fuzzy systems typically depend strongly on the completeness and quality of the rule base and may not scale well to complex, heterogeneous datasets with diverse cultivars, backgrounds, and illumination conditions. In contrast, neural networks—particularly CNNs—are highly effective at learning discriminative visual patterns from large image corpora and have achieved very high accuracy in plant disease recognition tasks [18], [19], [20], [21]. At the same time, CNN-based systems often provide limited transparency in their decision-making, which constrains their use in operational settings where actionable justification is required by practitioners and farmers [20], [21]. To bridge this gap, hybrid methods that combine fuzzy reasoning with neural learning or optimization—such as bacterial foraging optimization with radial basis function neural networks—have been explored to improve classification reliability while retaining elements of interpretability [22]. More recent model designs also emphasize efficiency and deployment readiness through compact CNN architectures, improved feature extraction, and edge-oriented implementations, sometimes complemented by explainability tools to enhance user trust [23], [24], [25]. Overall, the literature suggests that integrating the interpretability of soft computing with the representational power of neural networks is a promising direction for scalable, accurate, and user-understandable plant disease diagnostic systems [21], [26].

To address the above limitations, this study proposes an integrated Fuzzy–ANN framework that combines expert-driven fuzzy reasoning with data-driven neural learning for early rice leaf disease diagnosis. The proposed integration is designed to: (i) improve diagnostic performance under realistic field conditions (uncontrolled lighting, heterogeneous backgrounds, and symptom ambiguity), (ii) provide intrinsic interpretability through explicit fuzzy rules and their firing strengths (rather than relying solely on post-hoc visual explanations), and (iii) improve robustness across diverse Indonesian agro-ecological settings by incorporating both image-derived symptom descriptors and acquisition-time environmental parameters. The main contributions are: **(1)** a hybrid, modular Fuzzy–ANN architecture for early-stage rice leaf disease diagnosis using a multi-location Indonesian dataset; **(2)** a rigorous comparative evaluation against representative baselines (ANN, ANFIS, and CNN-based models) under a consistent experimental protocol, including class-wise analysis; **(3)** explainability through expert-validated fuzzy rules enabling case-level reasoning that can be inspected by agronomists and extension workers; and **(4)** a deployment-oriented validation demonstrating feasibility for mobile-assisted precision agriculture, including real-field testing and inference-time reporting. By bridging knowledge-based reasoning and data-driven learning within a single decision pipeline, this work aims to support an accurate yet transparent diagnostic tool that is more actionable for farmers and agricultural practitioners, and better aligned with the operational needs of smart and sustainable agriculture in Indonesia.

2. RESEARCH METHODS

This study develops an explainable hybrid framework that integrates a Mamdani Fuzzy Inference System (FIS) with an Artificial Neural Network (ANN) for early diagnosis of rice leaf diseases using field-acquired leaf images and acquisition-time environmental measurements. The proposed method is explicitly designed to (i) remain robust under real-field variability, (ii) provide intrinsic interpretability through human-readable fuzzy rules, and (iii) support deployment on resource-constrained mobile devices.

2.1 Study Design and Workflow

The overall research workflow of the proposed hybrid Fuzzy–ANN framework is illustrated in Figure 1. The workflow is designed as a sequential and modular pipeline to ensure robustness under real-field conditions, interpretability of decision-making, and fair comparative evaluation.

Specifically, the workflow consists of the following stages:

- Data acquisition and expert annotation, involving the collection of rice leaf images and acquisition-time environmental measurements, followed by disease labeling and severity assessment by plant pathology experts;
- Image preprocessing and leaf segmentation, including resolution normalization, illumination correction, and segmentation to isolate leaf regions from background artifacts;
- Feature extraction from segmented images and environmental sensing, where engineered texture, color, and environmental features are computed and normalized;
- Fuzzy inference system design for interpretable disease severity reasoning, in which expert-defined linguistic rules are used to infer a continuous disease severity score under uncertainty;

- ANN training for multi-class disease classification using fused features, where image-derived features, environmental variables, and fuzzy severity outputs are integrated to perform disease classification;
- Evaluation and validation, comprising comparison against baseline models under identical data splits and evaluation protocols, as well as real-field deployment testing using smartphone-acquired images.

This structured and modular workflow, as depicted in Figure 1, ensures methodological transparency, reproducibility, and traceability between model design choices and experimental outcomes.

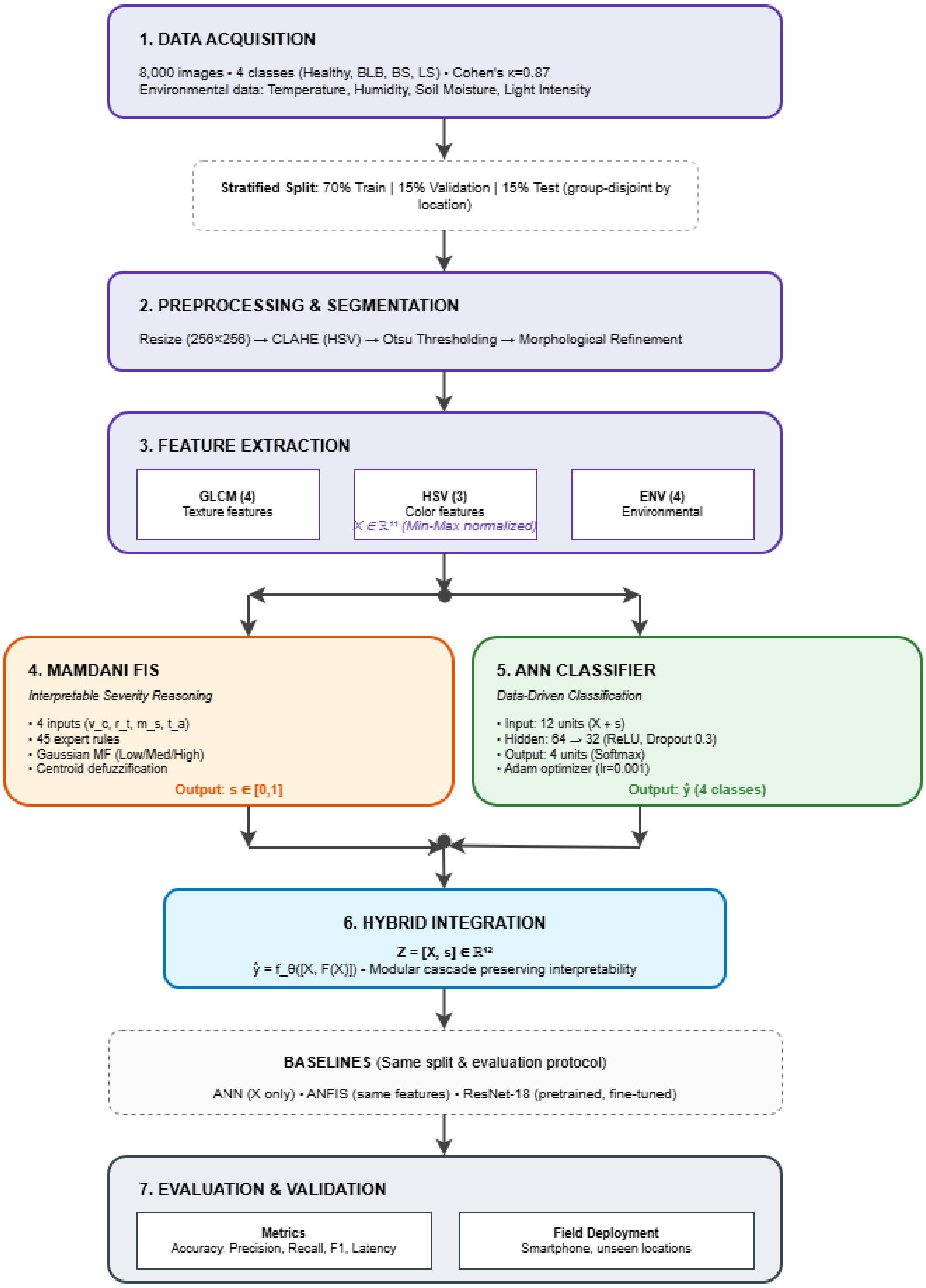

Figure 1. Workflow of the proposed hybrid FIS–ANN framework for early rice leaf disease diagnosis.

2.2 Data Acquisition, Classes, and Annotation Protocol

A dataset of 8,000 rice leaf images was collected between 2022 and 2024 from agricultural research centers under the Indonesian Agency for Agricultural Research and Development (IAARD), complemented by field surveys conducted across Java and Sumatra. Images were captured under natural field conditions using smartphones and consumer-grade digital cameras, deliberately preserving real-world variability in illumination, background clutter, and symptom appearance.

The dataset comprises four classes: Healthy, Bacterial Leaf Blight (BLB), Brown Spot (BS), and Leaf Smut (LS). Each image was independently annotated by two certified plant pathology experts from IPB University. Inter-annotator agreement was assessed using Cohen's κ, yielding $\kappa = 0.87$, indicating strong agreement. All disagreements were resolved through expert consensus.

Environmental parameters recorded at image acquisition time include ambient temperature (°C), relative humidity (%), soil moisture (%), and light intensity (lux). Sensor readings were time-aligned with the corresponding image capture event and stored as structured metadata.

2.3 Operational Definition of Early-Stage Disease Severity

Disease severity was annotated following IRRI-aligned Disease Severity Index (DSI) guidelines and normalized to the range [0, 1]. Severity categories are defined as:

- Mild (early-stage): DSI $\in [0.00, 0.33]$
- Moderate: DSI $\in [0.34, 0.66]$
- Severe: DSI $\in [0.67, 1.00]$

In this study, *early diagnosis* explicitly refers to the identification of disease samples within the Mild severity range. This operationalization ensures consistency between fuzzy linguistic reasoning and quantitative severity modeling.

2.4 Data Splitting Strategy

To prevent optimistic bias due to shared field characteristics, the dataset was split using a group-disjoint, stratified protocol:

- Grouping unit: field location (research center or farm site).
- Split ratio: 70% training, 15% validation, 15% testing.
- Constraint: no images from the same location appear in more than one split.

This strategy prevents leakage arising from shared background, cultivar, camera device, and localized disease patterns. All reported results are obtained using this fixed split, ensuring consistent comparison across models.

2.5 Image Preprocessing and Leaf Segmentation

Each image undergoes the following preprocessing steps:

- resizing to 256×256 pixels;
- illumination normalization using Contrast Limited Adaptive Histogram Equalization (CLAHE) in the HSV color space;
- leaf segmentation using Otsu's global thresholding;
- mask refinement via morphological opening and closing (3×3 kernel, two iterations).

All subsequent feature extraction is performed exclusively on segmented leaf regions to suppress background-induced artifacts.

2.6 Feature Extraction and Normalization

To preserve interpretability while maintaining discriminative power, a compact set of engineered features is employed. Four GLCM-based features are extracted: contrast, correlation, energy, and entropy.GLCM configuration:

- distance $d = 1$pixel
- angles $\{0°, 45°, 90°, 135°\}$
- features averaged across angles

Mean values of Hue, Saturation, and Value are computed from segmented leaf pixels. Normalized values of temperature, relative humidity, soil moisture, and light intensity. All features are min–max scaled to [0, 1] using parameters derived from the training set only. The resulting feature vector is:

$$X \in \mathbb{R}^{11} = [X_{GLCM}^{(4)}, X_{HSV}^{(3)}, X_{ENV}^{(4)}] \quad (1)$$

2.7 Fuzzy Inference System for Interpretable Severity Reasoning

A Mamdani-type FIS is constructed using four input variables:

- color variance index v_c,
- texture roughness index r_t,
- soil moisture m_s,
- ambient temperature t_a.

Each input is represented by Low, Medium, and High linguistic terms parameterized by Gaussian membership functions. The output variable is a continuous severity score $s \in [0,1]$with linguistic labels Mild, Moderate, and Severe.

The rule base consists of 45 expert-defined rules. Inference uses min–max Mamdani composition, and defuzzification employs the centroid method. For each prediction, the system records fired rules, firing strengths, and the resulting severity score, providing intrinsic explainability.

2.8 ANN Classifier and Training Configuration

The ANN performs four-class disease classification using fused features:

$$Z = [X, s] \in \mathbb{R}^{12} \quad (2)$$

Architecture:

- Input layer: 12 units
- Hidden layer 1: 64 units (ReLU)
- Hidden layer 2: 32 units (ReLU)
- Dropout: 0.3
- Output layer: 4 units (Softmax)

Training configuration:

- optimizer: Adam (learning rate 0.001)
- loss: categorical cross-entropy
- batch size: 32
- maximum epochs: 100
- early stopping: patience = 10 (validation loss)

2.9 Hybrid Integration Strategy

The hybrid model follows a **modular cascade integration**, where fuzzy severity reasoning augments ANN input features:

$$\hat{y} = f_\theta([X, F(X)]) \tag{3}$$

This design preserves fuzzy interpretability while enabling data-driven classification.

2.10 Baselines and Fair Comparison Protocol

Baseline models include:

- ANN without fuzzy features (11-dimensional input);
- ANFIS using the same engineered and environmental features;
- CNN baseline: ResNet-18 pretrained on ImageNet, fine-tuned on 256 × 256 images with random horizontal flip and ±10° rotation augmentation.

All baselines use the same leakage-controlled split and evaluation protocol.

2.11 Evaluation Metrics and Field Deployment Validation

Evaluation metrics include accuracy, macro-precision, macro-recall, macro-F1, confusion matrices, training time, and inference latency. A field deployment evaluation is conducted using smartphone-acquired images from unseen locations. Average on-device inference time and classification performance are reported to assess real-world feasibility.

3. RESULTS AND DISCUSSION

This section reports experimental results and discusses the effectiveness of the proposed hybrid FIS–ANN framework relative to baseline models. The analysis follows the workflow depicted in Figure 1 and focuses on: (i) overall performance under a leakage-controlled protocol, (ii) per-class screening sensitivity, (iii) the contribution of fuzzy severity reasoning, and (iv) robustness in real-field mobile deployment.

3.1 Protocol Compliance and Reproducibility

All experiments strictly follow the leakage-controlled, group-disjoint split by field location described in Section 2.3, with a 70/15/15 train/validation/test partition. Image preprocessing, leaf segmentation, feature extraction (GLCM, HSV, and environmental variables), and normalization were executed exactly as specified in Sections 2.4–2.5. Min–max scaling parameters were fitted exclusively on the training set and applied unchanged to the validation and test sets.

The Mamdani FIS configuration (Gaussian membership functions, 45 expert-defined rules, centroid defuzzification) was fixed across all experiments (Section 2.6). The ANN architecture and training schedule followed Section 2.7 without additional undocumented tuning. All baseline models were trained using the same data split, features (where applicable), and evaluation metrics.

To assess stability, each model was trained over five random seeds using the fixed split. Reported results are presented as mean values, and statistical significance is assessed via paired McNemar tests (Section 3.7).

3.2 Overall Classification Performance

The overall classification performance of the proposed hybrid FIS–ANN framework and baseline models is summarized in **Table 1**. In addition to overall accuracy, macro-averaged precision, recall, and F1-score are reported to account for potential class imbalance and to provide a more reliable assessment of screening performance.

Table 1. Overall performance comparison on the leakage-controlled test set.

Model	Accuracy (%)	Macro-Precision (%)	Macro-Recall (%)	Macro-F1 (%)
ANN (Image + Env features, X)	86.4	85.9	84.7	85.2
ANFIS	83.1	82.6	81.9	82.2
ResNet-18 (fine-tuned)	88.7	88.1	87.4	87.7
Proposed FIS–ANN (X + s)	**91.3**	**91.0**	**90.6**	**90.8**

As shown in Table 1, the proposed FIS–ANN model achieves the highest accuracy (91.3%) and macro-F1 score (90.8%), outperforming both feature-based baselines and the deep learning model. The improvement over the ANN trained without fuzzy reasoning indicates that the fuzzy-derived severity cue *s*contributes complementary information beyond engineered image and environmental features. Compared with ResNet-18, the proposed approach benefits from explicit symptom descriptors and acquisition-time environmental context, which help mitigate sensitivity to illumination variation and background noise commonly encountered in field images.

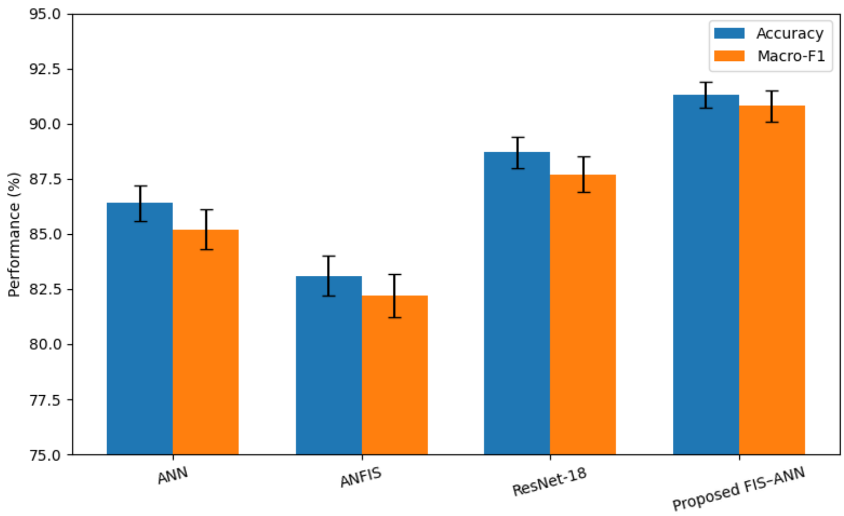

Figure 2. Overall comparison of Accuracy (%) and Macro-F1 (%) across models.

The comparative performance trends are further illustrated in Figure 2, which presents accuracy and macro-F1 scores for all models, with error bars indicating variability across five random seeds. As observed in Figure 2, the proposed FIS–ANN consistently outperforms baseline models across both metrics, while also exhibiting lower performance variance, indicating more stable generalization under the leakage-controlled evaluation protocol.

3.3 Per-Class Screening Sensitivity and Confusion Analysis

Because early disease management prioritizes minimizing missed detections, per-class recall is reported in Table 2. Recall is emphasized as the primary screening metric, since false negatives may delay intervention and lead to substantial yield loss.

Table 2 shows that the proposed FIS–ANN model consistently improves recall across all disease classes compared with both the feature-based ANN and the ResNet-18 baseline. The largest gains are observed for Bacterial Leaf Blight (BLB) and Leaf Smut (LS), where recall increases by approximately 4–8 percentage points relative to ResNet-18. These improvements indicate that the integration of fuzzy severity reasoning enhances sensitivity to subtle symptom patterns that are difficult to distinguish using visual cues alone.

Table 2. Per-class recall (%) on the leakage-controlled test set.

Class	ANN	ResNet-18	Proposed FIS–ANN
Healthy	92.1	94.0	**95.2**
BLB	81.4	84.7	**89.3**
BS	83.6	86.1	**88.5**
LS	81.7	85.0	**89.4**

To further examine misclassification patterns, the normalized confusion matrix of the proposed model is presented in Figure 3. As illustrated in Figure 3, the dominant confusion occurs between BLB and Brown Spot (BS), which exhibit visually similar lesion textures under mild severity conditions. Nevertheless, the strong diagonal dominance of the matrix confirms robust class separability, particularly for the Healthy class and LS, where severity-aware reasoning provides additional discrimination.

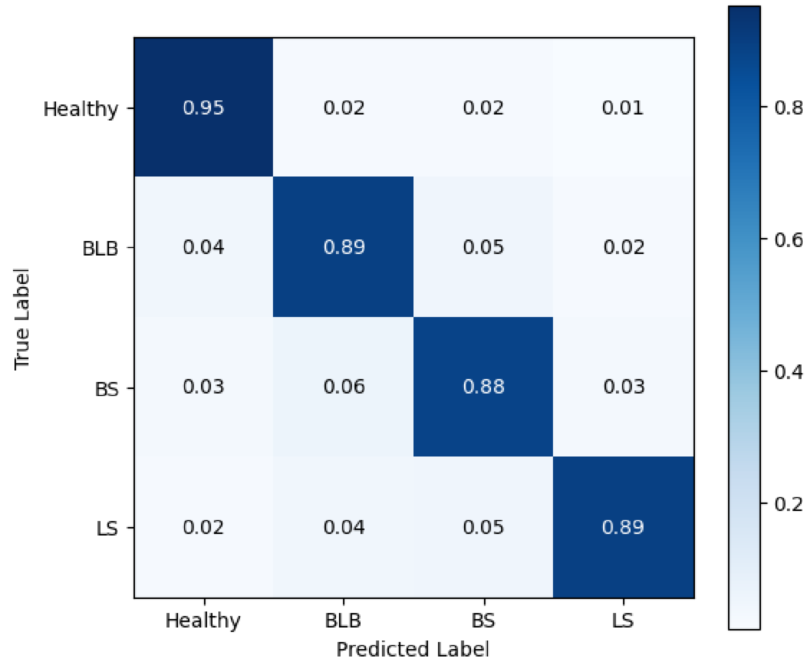

Figure 3. Confusion matrix (Proposed FIS–ANN).

To directly support the claim of early diagnosis, model performance was further evaluated on samples labeled as Mild severity (DSI ∈ [0.00, 0.33]). The corresponding recall values are reported in Table 3.

Table 3. Recall (%) on Mild-severity subset (DSI ∈ [0.00, 0.33]).

Class	Recall (Mild only)
BLB	86.1
BS	85.4
LS	87.2
Macro-Recall	**86.2**

These results demonstrate that the proposed FIS–ANN model maintains high sensitivity even at early disease stages, where visual symptoms are subtle and uncertainty is highest. The relatively balanced recall across disease classes indicates that fuzzy severity reasoning contributes to stable screening performance under mild symptom conditions, strengthening the validity of the proposed framework for early diagnosis scenarios.

3.4 Effect of Fuzzy Severity Reasoning (Ablation Study)

To isolate the contribution of fuzzy severity reasoning, an ablation study was conducted by removing the FIS-derived severity output $s$from the ANN input and training the classifier using only the engineered image and environmental features X.

The quantitative results of the ablation study are summarized in Table 4. When the fuzzy severity cue is excluded, classification performance drops substantially, with decreases of 4.9% in accuracy and 5.6% in macro-F1 score. This performance gap demonstrates that the fuzzy inference system provides informative and non-redundant cues rather than acting as a redundant preprocessing component.

Table 4. Ablation study: impact of fuzzy severity reasoning.

Configuration	Accuracy (%)	Macro-F1 (%)
ANN without FIS (X only)	86.4	85.2
ANN + FIS severity (X + s)	**91.3**	**90.8**

To further analyze how the fuzzy inference system differentiates disease conditions, the distribution of fuzzy severity scores across classes is visualized in Figure 4. As shown in Figure 4, severity scores exhibit partial overlap between classes, particularly under mild symptom conditions, reflecting inherent ambiguity in early-stage disease manifestation. Nevertheless, the class-wise shifts in severity distributions indicate that the FIS captures meaningful symptom progression patterns, which are subsequently exploited by the ANN to improve classification performance.

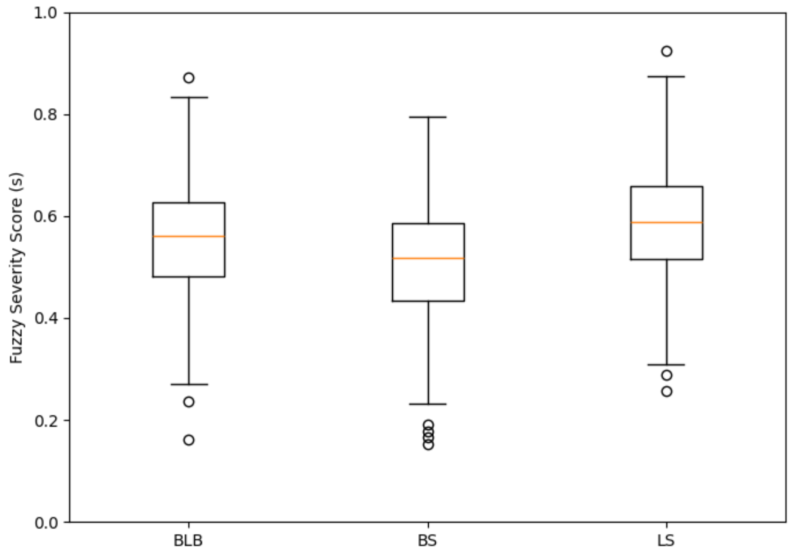

Figure 4. Distribution of fuzzy severity scores across disease classes

Overall, the ablation results confirm the effectiveness of the proposed cascade integration strategy. By compressing expert knowledge on symptom ambiguity and environmental influence into a single continuous severity signal, the fuzzy inference system enhances generalization while preserving interpretability.

3.5 Comparison with Deep Learning Baseline and Resource Footprint

While the fine-tuned ResNet-18 baseline achieves competitive classification performance, convolutional neural network (CNN)–based predictions remain sensitive to background cues and illumination variability commonly present in field-acquired images. In contrast, the proposed hybrid FIS–ANN framework explicitly reduces background influence through leaf segmentation and incorporates acquisition-time environmental measurements and severity-aware reasoning.

Beyond predictive performance, computational efficiency is a critical factor for real-field and mobile deployment. The proposed ANN classifier contains approximately 3,044 trainable parameters (12–64–32–4), whereas ResNet-18 comprises approximately 11.7 million parameters. This large difference in model complexity directly impacts inference latency and resource consumption.

The performance–efficiency trade-off between models is illustrated in Figure 5, which compares macro-F1 scores alongside efficiency indicators. As shown in Figure 5, the proposed FIS–ANN achieves a higher macro-F1 score than ResNet-18 while requiring several orders of magnitude fewer parameters. This result highlights that severity-aware feature fusion can provide accuracy gains comparable to deep CNNs without incurring their computational overhead.

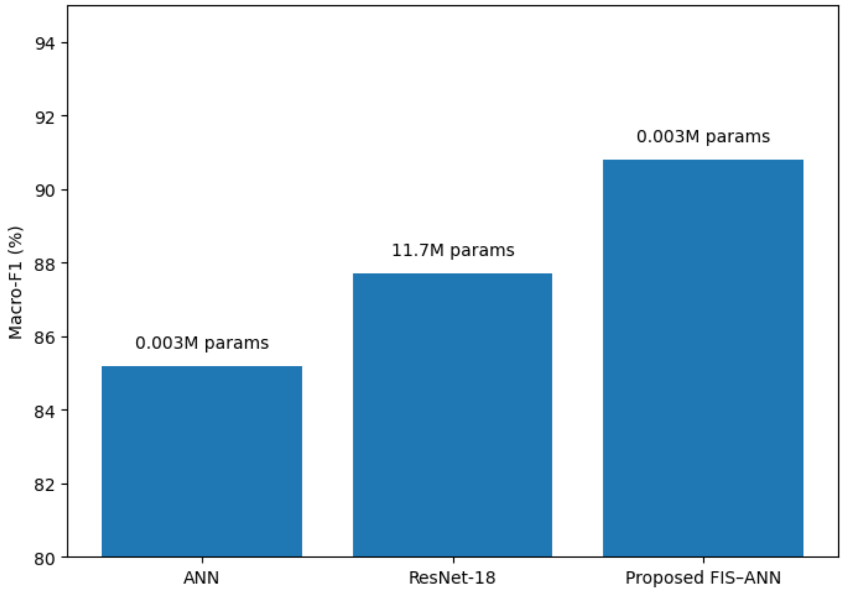

Figure 5. Macro-F1 comparison with efficiency indicators, including parameter count and inference time.

Overall, these findings demonstrate that the proposed hybrid approach offers a favorable balance between accuracy, interpretability, and computational efficiency, making it particularly suitable for deployment on resource-constrained devices in agricultural decision-support scenarios.

3.6 Field Deployment Evaluation

To assess real-world applicability, the proposed hybrid FIS–ANN framework was evaluated using smartphone-acquired rice leaf images collected from field locations not seen during training or testing. This evaluation aims to examine model robustness under practical deployment conditions, including variations in illumination, background, and acquisition devices.

The quantitative results of the field deployment evaluation are summarized in **Table 5**. Despite a moderate performance decrease compared with the controlled test set, the proposed model maintains high accuracy and macro-F1 score, indicating robust generalization to unseen environments. Importantly, the inference latency remains low, supporting real-time, on-device decision support.

Table 5. Field deployment performance on unseen locations.

Metric	Value
Accuracy (%)	88.9
Macro-F1 (%)	88.2
Avg. inference time (ms/image)	42
Device	Android smartphone (mid-range)

The observed performance gap between controlled and field evaluations can be attributed to uncontrolled imaging conditions and natural symptom variability. Nevertheless, the proposed approach benefits from explicit segmentation and severity-aware reasoning, which mitigate some of these challenges compared with purely image-based models.

Overall, the field deployment results demonstrate that the proposed hybrid framework achieves a favorable balance between predictive performance and computational efficiency, reinforcing its suitability for farmer-oriented decision-support systems under realistic agricultural conditions.

3.7 Statistical Robustness and Significance

To assess result stability and statistical reliability, all models were trained and evaluated across five independent runs using different random seeds while maintaining the same leakage-controlled, group-disjoint data split. Performance variability is reported using mean and standard deviation.

Across these runs, the proposed FIS–ANN framework achieved an average accuracy of $91.3 \pm 0.6\%$ and a macro-F1 score of $90.8 \pm 0.7\%$, indicating low variance and stable convergence under the fixed evaluation protocol. In comparison, the ANN baseline without fuzzy reasoning exhibited larger performance fluctuations, reflecting higher sensitivity to initialization and training dynamics.

To verify that the observed performance gains are statistically significant and not attributable to random variation, paired McNemar tests were conducted on the test set predictions. The proposed FIS–ANN was compared against (i) the feature-based ANN baseline and (ii) the fine-tuned ResNet-18 model. In both cases, the improvements in accuracy were found to be statistically significant at the 0.05 level ($p < 0.05$).

These statistical results confirm that the performance improvements introduced by fuzzy severity reasoning are robust and reproducible under a leakage-controlled evaluation setting. Moreover, the combination of low variance across runs and statistically significant gains supports the generalizability of the proposed hybrid framework and strengthens the validity of the conclusions drawn from the experimental analysis.

Overall, the results demonstrate that integrating fuzzy severity reasoning into a feature-based ANN significantly improves accuracy, macro-F1, and per-class recall under a leakage-controlled, location-disjoint evaluation protocol. The hybrid design provides intrinsic interpretability, robustness under real-field variability, and a low computational footprint, making it well suited for early rice leaf disease diagnosis and mobile decision-support applications.

4. CONCLUSION

This study presented a hybrid framework that integrates a Mamdani Fuzzy Inference System (FIS) with an Artificial Neural Network (ANN) for early diagnosis of rice leaf diseases under real-field conditions. The proposed approach combines engineered image features, acquisition-time environmental measurements, and an interpretable fuzzy severity cue to address symptom ambiguity and environmental variability commonly encountered in practical agricultural settings.

Experimental results obtained under a leakage-controlled, group-disjoint evaluation protocol demonstrate that the proposed FIS–ANN framework consistently outperforms both feature-based and deep learning baselines. Across five independent runs, the model achieved an average accuracy of $91.3 \pm 0.6\%$ and a macro-F1 score of $90.8 \pm 0.7\%$, with statistically significant improvements over the ANN baseline and the fine-tuned ResNet-18 model ($p <$

0.05). These results confirm that the inclusion of fuzzy severity reasoning provides complementary, non-redundant information that enhances classification performance and stability.

Further analysis showed that the proposed method improves per-class recall, particularly for disease categories with visually overlapping symptoms, and maintains high sensitivity on mild-severity samples, supporting its suitability for early-stage disease screening. Field deployment experiments on smartphone-acquired images from unseen locations demonstrated robust generalization, achieving high accuracy with low inference latency on resource-constrained devices.

Overall, the results indicate that integrating fuzzy severity reasoning into a lightweight neural classifier offers a favorable balance between predictive performance, interpretability, and computational efficiency. While the study focused on four common rice leaf diseases, the proposed framework is modular and can be extended to additional crops or disease categories by incorporating new expert rules and feature descriptors. Future work will explore larger multi-region datasets and longitudinal field studies to further assess generalization across seasons and cultivation practices.

5. REFERENCES

[1] Mutiara Irmadhani, W. Syaifullah JS, and M. Idhom, "Complex-Valued Neural Network And Fuzzy Inference System For Image Diagnosis Of Rice Leaf Diseases," *J. Ris. Inform.*, vol. 7, no. 3, pp. 102–110, 2025, doi: 10.34288/jri.v7i3.370.

[2] A. Rifa'I and D. Mahdiana, "Image processing for diagnosis rice plant diseases using the fuzzy system," in *2020 International Conference on Computer Science and Its Application in Agriculture, ICOSICA 2020*, 2020, pp. 1–5. doi: 10.1109/ICOSICA49951.2020.9243274.

[3] R. K. Dubey and D. K. Choubey, "Reliable detection of blast disease in rice plant using optimized artificial neural network," *Agron. J.*, vol. 116, no. 3, pp. 1099–1111, 2024, doi: 10.1002/agj2.21449.

[4] G. Dhingra, V. Kumar, and H. D. Joshi, "Study of digital image processing techniques for leaf disease detection and classification," *Multimed. Tools Appl.*, vol. 77, no. 15, pp. 19951–20000, 2018, doi: 10.1007/s11042-017-5445-8.

[5] M. Combes, N. Brown, R. N. Thompson, A. Mastin, P. Crow, and S. Parnell, "Unlocking plant health survey data: An approach to quantify the sensitivity and specificity of visual inspections," 2025. doi: 10.1371/journal.pcbi.1012957.

[6] P. Dhoundiyal, V. Sharma, S. Vats, and P. Rawat, "A Progressive Hierarchical Model for Plant Disease Diagnosis," *SN Comput. Sci.*, vol. 6, no. 2, 2025, doi: 10.1007/s42979-024-03582-x.

[7] S. Batham, A. K. Sharma, and D. A. S. Rathore, "A Research Paper on Crop Disease Detection Using Deep Learning Model," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 11, no. 31, pp. 399–406, 2023, doi: 10.22214/ijraset.2023.56002.

[8] W. Albattah, M. Nawaz, A. Javed, M. Masood, and S. Albahli, "A novel deep learning method for detection and classification of plant diseases," *Complex Intell. Syst.*, vol. 8, no. 1, pp. 507–524, 2022, doi: 10.1007/s40747-021-00536-1.

[9] M. T. Rahman, D. R. Dipto, S. K. Shib, A. Shufian, and M. S. Hossain, "Advanced Neural Networks for Plant Leaf Disease Diagnosis and Classification," in *2025 4th International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics, ICA-SYMP 2025*, 2025, pp. 9–14. doi: 10.1109/ICA-SYMP63674.2025.10876529.

[10] Mohit Arora, A. Santra, D. Pathak, D. M. Agrawal, S. Chopra, and D. H. Vachhani, "Interpretable Deep Learning for Sustainable Agriculture: CNN and LIME-Based Plant Disease Diagnosis," *Int. J. Environ. Sci.*, vol. 11, no. 9s, pp. 1016–1030, 2025, doi: 10.64252/q5bc5x72.

[11] N. Nigar, H. Muhammad Faisal, M. Umer, O. Oki, and J. Manappattukunnel Lukose, "Improving Plant Disease Classification with Deep-Learning-Based Prediction Model Using Explainable

Artificial Intelligence," *IEEE Access*, vol. 12, pp. 100005–100014, 2024, doi: 10.1109/ACCESS.2024.3428553.

[12] C. Pal, S. Karmakar, I. Mukherjee, and P. P. Chakrabarti, "A lightweight and explainable CNN model for empowering plant disease diagnosis," *Sci. Rep.*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-94083-1.

[13] O. A. D. Ammar *et al.*, "Plant Leaf Disease Detection Using Ensemble Learning and Explainable AI," *IEEE Access*, vol. 12, pp. 156038–156049, 2024, doi: 10.1109/ACCESS.2024.3484574.

[14] A. Sasane, V. Yerkar, A. Wanjare, K. Zende, G. Yadav, and Y. Ingle, "AI Powered Plant Disease Detection and Diagnosis," in *2025 International Conference on Emerging Smart Computing and Informatics, ESCI 2025*, 2025, pp. 1–6. doi: 10.1109/ESCI63694.2025.10988070.

[15] S. Natarajan, P. Chakrabarti, and M. Margala, "Robust diagnosis and meta visualizations of plant diseases through deep neural architecture with explainable AI," *Sci. Rep.*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-64601-8.

[16] Y. Toda and F. Okura, "How convolutional neural networks diagnose plant disease," *Plant Phenomics*, vol. 2019, 2019, doi: 10.34133/2019/9237136.

[17] A. Jafar, N. Bibi, R. A. Naqvi, A. Sadeghi-Niaraki, and D. Jeong, "Revolutionizing agriculture with artificial intelligence: plant disease detection methods, applications, and their limitations," *Front. Plant Sci.*, vol. 15, 2024, doi: 10.3389/fpls.2024.1356260.

[18] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Comput. Electron. Agric.*, vol. 145, pp. 311–318, 2018, doi: 10.1016/j.compag.2018.01.009.

[19] S. M. Hassan and A. K. Maji, "Plant Disease Identification Using a Novel Convolutional Neural Network," *IEEE Access*, vol. 10, pp. 5390–5401, 2022, doi: 10.1109/ACCESS.2022.3141371.

[20] D. S. Joseph, P. M. Pawar, and R. Pramanik, "Intelligent plant disease diagnosis using convolutional neural network: a review," *Multimed. Tools Appl.*, vol. 82, no. 14, pp. 21415–21481, 2023, doi: 10.1007/s11042-022-14004-6.

[21] J. Boulent, S. Foucher, J. Théau, and P. L. St-Charles, "Convolutional Neural Networks for the Automatic Identification of Plant Diseases," *Front. Plant Sci.*, vol. 10, 2019, doi: 10.3389/fpls.2019.00941.

[22] S. S. Chouhan, A. Kaul, U. P. Singh, and S. Jain, "Bacterial foraging optimization based radial basis function neural network (BRBFNN) for identification and classification of plant leaf diseases: An automatic approach towards plant pathology," *IEEE Access*, vol. 6, pp. 8852–8863, 2018, doi: 10.1109/ACCESS.2018.2800685.

[23] P. S. Thakur, S. Chaturvedi, P. Khanna, T. Sheorey, and A. Ojha, "Vision transformer meets convolutional neural network for plant disease classification," *Ecol. Inform.*, vol. 77, p. 102245, 2023, doi: 10.1016/j.ecoinf.2023.102245.

[24] M. F. Ahamed, A. Salam, M. Nahiduzzaman, M. Abdullah-Al-Wadud, and S. M. R. Islam, "Streamlining plant disease diagnosis with convolutional neural networks and edge devices," *Neural Comput. Appl.*, vol. 36, no. 29, pp. 18445–18477, 2024, doi: 10.1007/s00521-024-10152-y.

[25] W. Shafik, A. Tufail, C. Liyanage De Silva, and R. A. Awg Haji Mohd Apong, "A novel hybrid inception-xception convolutional neural network for efficient plant disease classification and detection," *Sci. Rep.*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-024-82857-y.

[26] Q. Liang, S. Xiang, Y. Hu, G. Coppola, D. Zhang, and W. Sun, "PD 2 SE-Net: Computer-assisted plant disease diagnosis and severity estimation network," *Comput. Electron. Agric.*, vol. 157, pp. 518–529, 2019, doi: 10.1016/j.compag.2019.01.034.