



ARTICLE

Fuzzy Attention Convolutional Neural Networks: A Novel Approach Combining Intuitionistic Fuzzy Sets and Deep Learning

Zheng Zhao¹, Doo Heon Song² and Kwang Baek Kim^{1,*}

¹Department of Artificial Intelligence, Silla University, Busan, 46958, Republic of Korea

²Department of Computer Games, Yong-In Art & Science University, Yongin, 17145, Republic of Korea

*Corresponding Author: Kwang Baek Kim. Email: gbkim@silla.ac.kr

Received: 29 September 2025; Accepted: 15 December 2025; Published: 12 March 2026

ABSTRACT: Deep learning attention mechanisms have achieved remarkable progress in computer vision, but still face limitations when handling images with ambiguous boundaries and uncertain feature representations. Conventional attention modules such as SE-Net, CBAM, ECA-Net, and CA adopt a deterministic paradigm, assigning fixed scalar weights to features without modeling ambiguity or confidence. To overcome these limitations, this paper proposes the Fuzzy Attention Network Layer (FANL), which integrates intuitionistic fuzzy set theory with convolutional neural networks to explicitly represent feature uncertainty through membership (μ), non-membership (ν), and hesitation (π) degrees. FANL consists of four core modules: (1) feature dimensionality reduction via global pooling, (2) fuzzy modeling using learnable clustering centers, (3) adaptive attention generation through weighted fusion of fuzzy components, and (4) feature refinement through residual connections. A cross-layer guidance mechanism is further introduced to enhance hierarchical feature propagation, allowing high-level semantic features to incorporate fine-grained texture information from shallow layers. Comprehensive experiments on three benchmark datasets—PathMNIST-30000, full PathMNIST, and BloodMNIST—demonstrate the effectiveness and generalizability of FANL. The model achieves $84.41 \pm 0.56\%$ accuracy and a 1.69% improvement over the baseline CNN while maintaining lightweight computational complexity. Ablation studies show that removing any component causes a 1.7%–2.0% performance drop, validating the synergistic contribution of each module. Furthermore, FANL provides superior uncertainty calibration (ECE = 0.0452) and interpretable selective prediction under uncertainty. Overall, FANL presents an efficient and uncertainty-aware attention framework that improves both accuracy and reliability, offering a promising direction for robust visual recognition under ambiguous or noisy conditions.

KEYWORDS: Attention mechanism; deep learning; intuitionistic fuzzy set; PathMNIST

1 Introduction

Deep convolutional neural networks (CNNs) have demonstrated exceptional performance in computer vision, with applications spanning image classification, object detection, semantic segmentation, and face recognition [1]. As network architectures have evolved toward greater depth to enhance representational capacity, challenges such as gradient degradation and information loss across layers have emerged [1]. To address these limitations, attention mechanisms have been extensively adopted in deep neural networks, enabling enhanced representation of salient features through the learning of adaptive importance weights [2,3].



Current mainstream attention mechanisms can be primarily categorized into three types: channel attention, spatial attention, and hybrid attention. Channel attention mechanisms (such as Squeeze-and-Excitation Networks, SE-Net [4]) focus on learning the importance differences among various channels; spatial attention mechanisms [5] concentrate on the salience of different spatial locations within feature maps; hybrid attention mechanisms such as Convolutional Block Attention Module (CBAM) [6] and Bottleneck Attention Module (BAM) [7] that attempt to simultaneously integrate information from both channel and spatial dimensions. These approaches effectively enhance the model's ability to focus on critical information through dynamic weighting of feature maps, significantly improving model performance and fully demonstrating the pivotal role of attention mechanisms in deep learning.

However, existing attention mechanisms still face three critical challenges. First, most methods adopt deterministic weight assignment strategies, using a single numerical value [8] to represent feature importance, which inadequately captures the diversity of feature importance across different semantic levels. Second, there is a lack of explicit modeling mechanisms for uncertainty features, leading to unstable performance when processing scenarios with blurred boundaries or noisy interference [9]. Finally, current mainstream attention mechanisms are predominantly limited to single-layer operations and fail to effectively exploit and integrate the complementarity of cross-layer attention information [6], thereby constraining the understanding capability for complex pathological structures. These limitations are particularly pronounced when dealing with medical imaging tasks that possess inherent diagnostic uncertainty, especially in pathological tissue classification tasks such as PathMNIST that require distinguishing subtle morphological differences. Traditional methods struggle to effectively handle transition regions between normal and pathological tissues, as well as similar features between different pathological types.

Fuzzy logic, as an important tool for handling uncertainty [10,11], particularly the intuitionistic fuzzy set theory proposed by Atanassov [12], provides a novel approach to addressing the aforementioned problems. Unlike traditional binary fuzzy sets, intuitionistic fuzzy sets characterize element attributes through a ternary representation of membership degree (μ), non-membership degree (ν), and hesitation degree (π), enabling more comprehensive expression of information uncertainty. This ternary structure conceptually aligns well with weight allocation in attention mechanisms: membership degree can be viewed as feature importance, non-membership degree reflects feature unimportance, while hesitation degree characterizes the uncertainty of feature weights. Recent studies have demonstrated that integrating fuzzy set theory with attention mechanisms can effectively enhance the continuity and robustness of deep models in medical image segmentation tasks [13]. However, how to efficiently integrate intuitionistic fuzzy set theory with attention mechanisms in deep learning remains a challenging open problem. The core difficulty lies in the fact that traditional fuzzy clustering algorithms typically exhibit high computational complexity, making them difficult to directly embed into the training and inference processes of deep neural networks.

To address these challenges, this paper proposes a novel Fuzzy Attention Network Layer (FANL) that organically integrates intuitionistic fuzzy set theory with attention mechanisms. FANL establishes membership relationships between features and predefined patterns through lightweight fuzzy clustering methods, utilizing the ternary representation of membership degree, non-membership degree, and hesitation degree to comprehensively characterize feature importance. Building upon this foundation, we develop an adaptive weight fusion strategy with learnable parameters that dynamically optimize the balance among different fuzzy measures across hierarchical feature representations. Simultaneously, an innovative cross-layer attention guide mechanism is introduced to effectively enhance feature representation capability. Compared to traditional attention mechanisms, FANL can explicitly model uncertainty in features while maintaining high computational efficiency and improved expressive capacity.

The main contributions of this paper can be summarized as follows:

1. A novel Fuzzy Attention Network Layer (FANL) is proposed that introduces intuitionistic fuzzy set theory into attention mechanisms in deep learning, utilizing a ternary fuzzy representation consisting of membership degree, non-membership degree, and hesitation degree to comprehensively characterize feature importance and uncertainty.
2. A lightweight fuzzy clustering strategy is designed for employing predefined clustering centers to replace traditional iterative optimization processes, effectively reducing the computational complexity of fuzzy modeling in a deep neural network structure and improving the model efficiency.
3. A learnable fuzzy weight fusion mechanism is proposed that automatically balances the contributions of ternary fuzzy metrics through adaptive parameters, enabling the model to effectively capture layer-specific uncertainty characteristics.
4. An innovative cross-layer attention guidance mechanism enables shallow-layer attention information to guide and enhance deep-layer feature representation, improving the representational capacity and contextual consistency of deep networks.
5. Comprehensive experimental validation was performed on the PathMNIST pathological image dataset [14] to evaluate the proposed FANL method. In our experiments, FANL exhibits better performance compared to state-of-the-art attention mechanisms across three key metrics: classification accuracy, model robustness, and interpretability. These findings demonstrate the promising application potential and strong generalizability of the proposed approach for pathological image analysis tasks.

The remainder of this paper is organized as follows: [Section 2](#) reviews related work covering attention mechanisms, fuzzy logic, and their applications in deep learning; [Section 3](#) presents the theoretical foundation and architectural design of the proposed FANL model; [Sections 4.1–4.3](#) describe the experimental setup, including datasets, evaluation metrics, and comparative methods, followed by systematic analysis of experimental results; [Section 4.4](#) discusses model advantages, limitations, and potential extension directions; [Section 5](#) concludes the paper and outlines future research directions.

2 Related Work

This section reviews research work closely related to the proposed FANL model, covering the following four aspects: attention mechanisms in deep learning, fusion methods of fuzzy logic and deep learning, uncertainty modeling techniques in deep neural networks, and cross-layer feature fusion with attention information propagation strategies.

2.1 Attention Mechanisms

Attention mechanisms enhance the modeling capability of deep neural networks by selectively focusing on critical information in input data and have become one of the core components in deep learning models. In the field of computer vision, attention mechanisms are typically categorized into three types: channel attention, spatial attention, and hybrid attention, which weight feature maps from different dimensions to enhance the model's capability to represent key features.

Channel attention mechanisms focus on modeling the importance differences among various channels in feature maps, serving as a key method for improving feature selection capability in deep networks. The Squeeze-and-Excitation Networks (SE-Net) proposed by Hu et al. [4] represents pioneering work in this direction. This method compresses spatial information through global average pooling and then utilizes a two-layer fully connected network to assign weights to each channel, thereby achieving channel-level feature recalibration. ECA-Net, proposed by Wang et al. [15], further optimizes the SE-Net structure by

replacing fully connected layers with one-dimensional convolution, significantly reducing parameters and computational overhead while maintaining good performance. Recent advances have continued to enhance channel attention mechanisms. Wang et al. [16] proposed an attention mechanism module that addresses the interaction between spatial and channel information, improving feature representation capability. Additionally, Xu et al. [17] introduced Sequential Fusion Attention (SFA) that efficiently combines spatial and channel dimensions while maintaining computational efficiency. Although these recent methods have achieved significant improvements in model performance, they generally rely on average or maximum values of channel features as the basis for attention weights, making it difficult to characterize potential uncertainty and ambiguity between features, thereby limiting model robustness and expressive capability in complex scenarios.

Spatial attention mechanisms aim to model the importance of different spatial positions in feature maps, thereby improving the model's perception capability for local and global spatial information. Non-Local Neural Networks proposed by Wang et al. [5] construct long-range dependencies, enabling each position to establish direct connections with any other position in the feature map, effectively capturing global spatial context. Spatial Transformer Networks proposed by Jaderberg et al. [18] achieve spatial invariance modeling for geometric transformations such as translation and scaling through explicit learning of spatial transformation parameters. More recently, a number of efficient spatial attention variants have been developed. For example, Hong et al. [19] proposed the Fast Non-Local Attention Network (FNLANet) to accelerate global context modeling. However, existing attention mechanisms typically operate on deterministic feature representations and lack explicit modeling of uncertainty—a critical limitation when dealing with fuzzy tissue boundaries, imaging noise, and ambiguous pathological patterns common in medical imaging [20]. This gap motivates the integration of fuzzy logic principles with attention mechanisms to enable explicit uncertainty quantification in feature representation.

Hybrid attention mechanisms aim to simultaneously model information from both channel and spatial dimensions, achieving more comprehensive feature selection and enhancement. CBAM (Convolutional Block Attention Module), proposed by Woo et al. [6], progressively strengthens feature representation by sequentially applying channel attention and spatial attention modules. Park et al. [7] adopted a parallel structure in BAM (Bottleneck Attention Module), simultaneously computing channel and spatial attention to improve modeling efficiency. Coordinate Attention proposed by Hou et al. [21] embeds positional information into channel attention through coordinate decomposition of spatial dimensions, achieving effective spatial information fusion while maintaining low computational cost. Although these hybrid attention mechanisms comprehensively utilize information from different dimensions and improve model performance, they generally adopt deterministic weight calculation methods and still fail to explicitly model uncertainty in feature importance, potentially resulting in insufficient expressive capability when processing ambiguous and complex scenarios.

Fuzzy-attention mechanisms for uncertainty modeling in deep learning. Rajafillah et al. [22] proposed Intuitionistic Fuzzy Pooling (INT-FUP), incorporating membership, non-membership, and hesitation degrees into CNN pooling layers to handle local imprecision in feature maps. Pencheva introduced Intuitionistic Fuzzy Deep Neural Network (IFDNN) [23], combining intuitionistic fuzzy sets with multilayer perceptrons by applying fuzzy logic to network weights for biomedical data classification. At the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) 2024 [24], a fuzzy-attention based border rendering network was presented for lung segmentation, demonstrating the utility of fuzzy logic in delineating uncertain anatomical boundaries. However, these works primarily focus on isolated components without end-to-end learnable frameworks or uncertainty calibration

metrics. FANL distinguishes itself through adaptive learnable fuzzy-attention with rigorous uncertainty quantification (Expected Calibration Error = 0.0452(ECE)) and hierarchical cross-layer guidance.

Existing attention mechanisms generally suffer from two core limitations. First, they typically adopt deterministic weight allocation strategies, using single numerical values to characterize feature importance, making it difficult to distinguish between “clearly unimportant” and “uncertain whether important”—two fundamentally different states, thereby limiting the model’s expressive capability for uncertain information. Second, most attention mechanisms operate only on single network layers, lacking modeling and fusion of cross-layer attention information, failing to fully exploit the complementarity between features at different levels, potentially leading to insufficient information utilization in complex tasks.

2.2 Integration of Fuzzy Logic and Deep Learning

Fuzzy logic, as an effective tool for handling uncertainty and ambiguity, has gradually attracted widespread attention from researchers regarding its integration with deep learning in recent years.

Fuzzy neural networks represent one of the earlier attempts to introduce fuzzy logic into neural network models. Lin and Lee [25] proposed a framework combining neural networks with fuzzy rules, implementing rule-based fuzzy inference mechanisms. Subsequently, Nauck et al. [26] developed the neuro-fuzzy classification system, further exploring the application potential of neuro-fuzzy classifiers in pattern recognition tasks. However, these models are predominantly based on shallow network architectures, lacking hierarchical representation capability for complex features, thus making it difficult to adapt to the modeling requirements of current deep learning architectures for large-scale data and complex tasks.

Application of fuzzy clustering in feature extraction represents another important direction in fuzzy logic research. Havens et al. [27] explored how to extend the Fuzzy C-Means (FCM) algorithm for large-scale data environments, improving its adaptability in practical tasks. Recent comprehensive reviews have demonstrated the effectiveness of intuitionistic fuzzy set theory and its extensions in medical image processing, particularly for handling uncertainty in pathological tissue analysis [28]. However, traditional fuzzy clustering algorithms typically rely on iterative optimization processes with high computational overhead, making them difficult to directly integrate into end-to-end deep neural network frameworks.

Meanwhile, deep attention mechanisms have shown remarkable success in medical image segmentation tasks. Several studies have demonstrated that attention-aware architectures can effectively capture both local and global contextual information in biomedical images [29]. Although these studies reveal the application prospects of attention mechanisms in medical imaging, systematic integration of “fuzzy logic and attention mechanisms” remains relatively scarce, particularly in pathological image classification, where effective frameworks and mechanism designs for fusing both approaches are still lacking.

2.3 Uncertainty Modeling in Deep Learning

In recent years, uncertainty modeling in deep learning has gradually become a research hotspot, primarily encompassing Bayesian methods, evidential theory methods, and fuzzy methods.

Bayesian deep learning explicitly represents uncertainty in model predictions through probabilistic modeling frameworks, possessing rigorous theoretical foundations and broad application prospects. Gal and Ghahramani [30] interpreted Dropout techniques as a Bayesian approximation method, introducing efficient Bayesian inference mechanisms into mainstream neural networks for the first time, providing a practical and feasible approach for estimating model uncertainty. Subsequently, Kendall and Gal [31] further distinguished between aleatory uncertainty and epistemic uncertainty, providing more refined semantic classification and modeling strategies for uncertainty modeling. Although Bayesian methods are theoretically complete

and highly interpretable, they still face challenges of high computational complexity and implementation difficulty in practical applications, limiting their widespread deployment in large-scale deep neural networks.

Evidential theory methods construct uncertainty representation based on Dempster-Shafer evidence theory, characterizing the credibility of model predictions through explicit modeling of belief, disbelief, and uncertainty components. The Evidential Deep Learning framework proposed by Sensoy et al. [32] utilizes Dirichlet distributions to represent evidential support for classification probabilities, achieving uncertainty quantification in classification tasks. Amini et al. [33] extended this framework to regression tasks, further broadening its application scope. These methods can distinguish different types of uncertainty (such as epistemic uncertainty and data uncertainty) but typically require significant modifications to loss functions and training procedures, increasing the complexity of model implementation and optimization.

Fuzzy set-based methods handle information uncertainty and ambiguity through fuzzy set theory, gradually attracting attention from deep learning researchers in recent years. The intuitionistic fuzzy set theory proposed by Atanassov [12] introduces a ternary representation of membership degree, non-membership degree, and hesitation degree based on traditional fuzzy sets, providing richer and more refined expressive capability for uncertainty modeling. Xu [34] conducted a systematic review of the development trajectory of intuitionistic fuzzy sets, organizing their theoretical foundations and main research directions. Recent advances have demonstrated the growing integration of fuzzy sets with deep learning architectures, with Zheng et al. [35] providing a comprehensive survey of fuzzy deep learning applications for uncertain medical data, highlighting the effectiveness of fuzzy-based approaches in handling imprecise and vague information in complex scenarios.

As an important unsupervised learning tool, **fuzzy clustering** also occupies an important position in uncertainty modeling. The Fuzzy C-Means algorithm (FCM) proposed by Bezdek [36] established the theoretical foundation of fuzzy clustering, with the core idea of allowing samples to belong to multiple categories with different membership degrees. The Intuitionistic Fuzzy C-Means algorithm (IFCM) proposed by Chaira [37] introduces non-membership degree and hesitation degree based on FCM, enhancing the expressive capability and flexibility of clustering. Although these methods perform excellently in traditional machine learning scenarios, their heavy reliance on iterative optimization with high computational overhead makes it difficult to directly embed them into end-to-end deep neural networks.

Fuzzy logic, particularly intuitionistic fuzzy set theory, has been widely applied in traditional pattern recognition and machine learning fields [38–40]. However, research combining intuitionistic fuzzy set theory with attention mechanisms in deep learning is still in its infancy, particularly in the field of computer vision.

Despite the success of attention mechanisms in computer vision, their application to medical image analysis reveals three fundamental limitations.

First, conventional mechanisms (SE-Net, CBAM, ECA-Net, CA) operate under a deterministic paradigm, assigning a single scalar weight $w \in [0, 1]$ to each feature. This binary encoding fails to capture the inherent ambiguity in medical images, where tumor infiltration zones, inflammatory regions, and tissue transitions exhibit fuzzy boundaries. A single confidence score (e.g., $w = 0.65$) conflates distinct scenarios: definite moderate importance, ambiguous features between states, and features on decision boundaries with high uncertainty. Second, these mechanisms lack explicit uncertainty quantification. SE-Net computes $w = \sigma(FC(GAP(x)))$ as a point estimate without modeling confidence in this estimate. In medical diagnosis, uncertain information is crucial for risk assessment (flagging ambiguous cases), selective prediction (abstaining from high-risk decisions), and interpretability (explaining why the model is uncertain). The absence of uncertainty modeling limits clinical applicability. Third, conventional approaches assume non-membership as $1-w$, failing to distinguish definite irrelevance (clear background) from unknown

relevance (ambiguous artifacts). For instance, a background region and an ambiguous artifact may both receive $w = 0.2$, despite requiring different handling strategies—suppression vs. cautious consideration.

These limitations motivate FANL's design: intuitionistic fuzzy sets provide a three-valued representation (μ, ν, π) that explicitly models feature ambiguity, quantifies uncertainty via hesitation degree π , and independently represents positive and negative evidence.

2.4 Uncertainty Quantification in Medical Imaging

Uncertainty quantification (UQ) is essential for deploying deep learning models in medical imaging applications, as it enables the identification of unreliable predictions and improves model trustworthiness [41]. In medical imaging, uncertainty arises from multiple sources, including image quality variations, acquisition artifacts, inter-observer variability, and inherent diagnostic ambiguity [42].

Existing UQ approaches can be categorized into several types. Bayesian deep learning methods have been widely explored for uncertainty estimation. Gal and Ghahramani [30] proposed Monte Carlo Dropout as a practical approximation to Bayesian inference, while Lakshminarayanan et al. [43] introduced deep ensembles that train multiple models to capture epistemic uncertainty. However, these methods require 5–20 forward passes per prediction, substantially increasing computational cost and limiting real-time applications. Post-hoc calibration methods provide an alternative approach. Guo et al. [44] demonstrated that temperature scaling effectively adjusts model outputs to align predicted confidence with empirical accuracy. While computationally efficient, these approaches operate on final predictions without modifying the feature learning process, potentially missing uncertainty information in intermediate representations.

FANL addresses these limitations by integrating uncertainty modeling directly into the feature attention mechanism through fuzzy logic. Unlike Bayesian methods requiring multiple inferences, FANL computes the hesitation degree π during single-pass forward propagation without additional overhead. Our calibration analysis demonstrates that FANL achieves ECE equals 0.0452, comparable to calibrated baselines, while the strong correlation between π and prediction errors (AUC equals 0.742) provides interpretable uncertainty signals. Furthermore, FANL's selective prediction framework achieves accuracy improvements from 84.96% to 92.81% at 30% coverage, demonstrating effective uncertainty-aware decision making with computational efficiency.

2.5 Cross-Layer Feature Fusion and Attention Propagation

Cross-layer information flow in deep neural networks is important for enhancing model feature representation capability and performance. Existing methods can be primarily categorized into two types: feature pyramid and multi-scale feature fusion methods, and cross-layer attention propagation mechanisms.

Feature pyramid and multi-scale feature fusion structures are dedicated to integrating features from different resolutions to enhance the model's perception capability for multi-scale targets. Feature Pyramid Networks (FPN) proposed by Lin et al. [45] construct top-down feature enhancement pathways and fuse semantic and spatial information from different levels through lateral connections, achieving efficient multi-scale representation. Based on this work, Liu et al. [46] proposed the Path Aggregation Network (PANet), introducing bottom-up information flow channels to further strengthen the depth and breadth of feature fusion. Although these methods achieve effective integration of cross-layer information at the feature level, they primarily focus on merging and enhancing multi-scale feature maps, while paying less attention to the propagation and collaborative modeling of attention information between different network layers, failing to fully exploit the potential value of inter-layer attention relationships in feature representation.

Attention guidance and propagation mechanisms are dedicated to propagating attention information between different levels of neural networks to achieve efficient collaboration and guidance of cross-layer features. Li et al. [47] explored strategies for using high-level semantic attention to guide low-level feature extraction, enhancing the semantic representation capability of bottom-layer features. The Attention Transfer method proposed by Zagoruyko and Komodakis [48] is primarily applied to knowledge distillation, achieving model compression and performance preservation by transferring attention distributions from teacher models to student models. Although these methods achieve cross-layer attention propagation to some extent, they generally focus on unidirectional guidance from high to low layers, or they are limited to specific tasks such as model compression. Research on how low-layer attention can reversely guide high-level semantic modeling is relatively scarce, particularly in contexts involving uncertainty modeling, where systematic mechanism design and theoretical support are still lacking.

In summary, although existing research has made considerable progress in attention mechanisms, fusion of fuzzy logic and deep learning, uncertainty modeling, and cross-layer feature interaction, the following challenges remain:

1. Existing attention mechanisms generally adopt deterministic weight modeling, lacking explicit representation of feature importance uncertainty.
2. Traditional fuzzy clustering algorithms have high computational complexity, making them difficult to directly embed into deep learning frameworks for end-to-end optimization.
3. There is a lack of methods that effectively integrate the ternary representation of intuitionistic fuzzy sets (membership degree, non-membership degree, and hesitation degree) with deep attention mechanisms.
4. There is an absence of low-to-high-level attention guidance mechanisms oriented toward fuzzy modeling contexts, limiting cross-layer propagation and representation of feature information.

To address these challenges, this paper proposes a novel Fuzzy Attention Network Layer (FANL) that organically integrates intuitionistic fuzzy set theory with attention mechanisms, featuring adaptive weight generation, efficient fuzzy modeling, and cross-layer attention guidance to provide new insights for uncertainty modeling in deep learning.

3 Fuzzy Attention Network Layer (FANL)

3.1 FANL Architecture

Traditional attention mechanisms typically enhance critical information in input features through deterministic weight allocation. Although they have achieved significant results in numerous vision tasks, they exhibit limitations in modeling capability when facing data scenarios with inherent ambiguity and uncertainty. In this paper, we propose the Fuzzy Attention Network Layer (FANL) provides a ternary fuzzy representation consisting of membership degree, non-membership degree, and hesitation degree to comprehensively model feature importance and explicitly express its uncertainty, as detailed in Figs. 1 and 2 in Section 4.1.2.

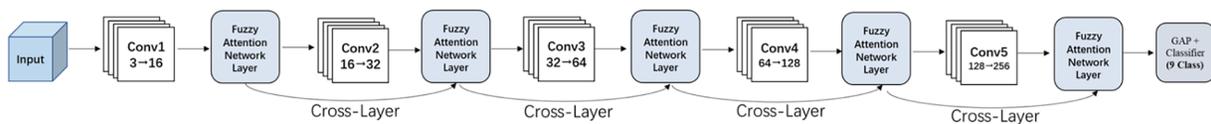


Figure 1: Overall architecture of the FANL-CNN model

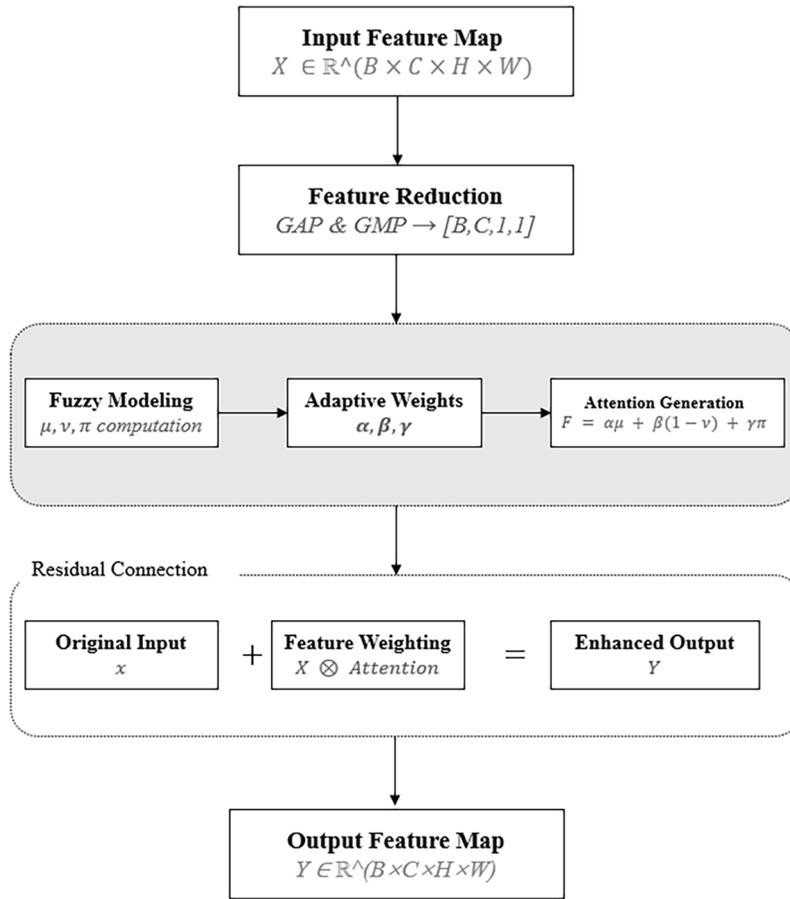


Figure 2: Architecture of the fuzzy attention network layer

The overall architecture of FANL comprises the following four core modules:

1. **Feature Dimensionality Reduction Module:** Extracts channel descriptors and compresses high-dimensional features to preserve semantic information.
2. **Fuzzy Modeling Module:** Computes membership degree, non-membership degree, and hesitation degree for each channel, implementing feature uncertainty modeling based on intuitionistic fuzzy sets.
3. **Attention Generation Module:** Fuses ternary fuzzy measures into attention weights using adaptively learned parameters that automatically optimize feature representation at different network depths.
4. **Feature Weighting Module:** Applies generated attention weights to enhance original features through weighting and introduces residual connections to preserve original information.

Additionally, FANL introduces a cross-layer attention guidance mechanism that enhances contextual consistency and robustness of high-level feature representation by guiding deep networks to explicitly utilize shallow-layer attention information.

3.2 Intuitionistic Fuzzy Modeling Module

3.2.1 Intuitionistic Fuzzy Sets

Intuitionistic Fuzzy Sets (IFS) were first proposed by Atanassov [12]. Unlike traditional fuzzy sets that model uncertainty solely through membership degree, intuitionistic fuzzy sets jointly characterize the

uncertainty state of elements through a ternary representation of membership degree (μ), non-membership degree (ν), and hesitation degree (π), providing richer and more refined uncertainty expression.

For any element x in the feature space, an intuitionistic fuzzy set A can be defined as:

$$A = \{ \langle x, \mu_A(x), \nu_A(x) \rangle \mid x \in X \}, \quad (1)$$

where:

- $\mu_A(x) \in [0, 1]$: represents the membership degree, i.e., the degree to which element x belongs to set A .
- $\nu_A(x) \in [0, 1]$: represents the non-membership degree, i.e., the degree to which element x does not belong to set A .
- $\pi_A(x) = 1 - \mu_A(x) - \nu_A(x)$: represents the hesitation degree, reflecting the uncertainty about the membership status of element x .

subject to the constraint condition:

$$0 \leq \mu_A(x) + \nu_A(x) \leq 1, \quad (2)$$

this ternary representation structure enables intuitionistic fuzzy sets to provide more expressive uncertainty modeling approaches when dealing with ambiguous, contradictory, or information-insufficient data.

3.2.2 Lightweight Fuzzy Clustering Strategy

Traditional Intuitionistic Fuzzy C-Means (IFCM) algorithms rely on iterative optimization processes with high computational cost, making it difficult to directly integrate into end-to-end training frameworks of deep learning. To meet the requirements of deep models for efficiency and differentiability, this paper proposes a lightweight fuzzy modeling method based on predefined clustering centers that significantly reduce computational complexity while preserving the ternary representation capability of intuitionistic fuzzy sets, thereby achieving efficient integration with neural network architectures.

Given input feature map $X \in \mathbb{R}^{B \times C \times H \times W}$, channel descriptors are first extracted through Global Average Pooling (GAP) and Global Max Pooling (GMP):

$$\mathbf{Z}_{avg} = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W \mathbf{X}, :, i, j, \quad (3)$$

$$\mathbf{Z}_{max} = \max_{i,j} \mathbf{X}, :, i, j, \quad (4)$$

for a set of K predefined clustering centers $\{C_1, C_2, \dots, C_K\}$, the distances from channel descriptors $\mathbf{Z} \in \mathbb{R}^d$ to each clustering center are computed to measure their similarity in fuzzy space. Euclidean distance is typically used for measurement, defined as follows:

$$d_{i,k} = \| \mathbf{Z}_i - C_k \|_2 \quad (5)$$

3.2.3 Adaptive Fuzzy Membership Computation

To enhance the model's adaptability to different data distributions, this paper proposes an adaptive sigma strategy. Traditional fuzzy modeling methods typically adopt fixed sigma values as scale parameters, and such static settings are difficult to adapt to diverse input feature distributions, potentially leading to distorted membership degree computations. Therefore, this paper designs a dynamic sigma generation mechanism through a learnable Multi-Layer Perceptron (MLP) that processes channel-wise feature

descriptors, making fuzzy measures more context-aware and responsive to feature characteristics as follows.

$$\sigma_{adaptive} = f_{\sigma}(Z) = 0.1 + 0.4 \cdot \text{Sigmoid}(W_2 \cdot \text{ReLU}(W_1 \cdot Z + b1) + b2), \quad (6)$$

where $Z \in \mathbb{R}^C$ denotes the channel descriptor, $W_1 \in \mathbb{R}^{\frac{C}{4} \times C}$ and $W_2 \in \mathbb{R}^{\frac{C}{4} \times 1}$ are learnable weight matrices, and $b1, b2$ are bias terms. The Sigmoid activation constrains the output to $[0, 1]$, which is then scaled to $[0.1, 0.5]$ to ensure numerical stability in Gaussian membership computation (Eq. (7)).

After obtaining the adaptive σ , this paper employs a Gaussian membership function to compute the membership degree for the distance $d_{i,k}$ between channel descriptor Z_i and the K clustering center C_K , defined as follows:

$$\mu_{i,k} = \exp\left(-\frac{d_{i,k}^2}{2\sigma_{adaptive}^2}\right), \quad (7)$$

where $\mu_{i,k}$ represents the membership degree of the i_{th} channel descriptor to the k_{th} clustering center, ranging from 0 to 1. $d_{i,k}$ denotes the Euclidean distance between the i_{th} channel descriptor Z_i and the k_{th} clustering center C_K , computed as $d_{i,k} = |x_i - C_K|$. $\sigma_{adaptive}$ is the adaptive scaling parameter derived from Eq. (6), which dynamically controls the width of the Gaussian membership function based on learned channel characteristics.

To ensure that the sum of membership degrees equals to 1, the normalization process is performed:

$$\mu_{i,k} = \frac{\mu_{i,k}}{\sum_{j=1}^K \mu_{i,j}}, \quad (8)$$

where $\mu_{i,k}$ represents the normalized membership degree.

Subsequently, based on the membership degree $\mu_{i,k}$, the corresponding non-membership degree $v_{i,k}$ is calculated using the following nonlinear transformation:

$$v_{i,k} = (1 - \mu_{i,k})^a, \quad (9)$$

where the hyperparameter $a = 1.6$ serves as a non-linearity factor, used to control the sensitivity relationship between membership degree and non-membership degree.

Finally, the hesitation degree $\pi_{i,k}$ is defined as the difference between 1 and the sum of membership degree and non-membership degree:

$$\pi_{i,k} = \max(0, 1 - \mu_{i,k} - v_{i,k}) \quad (10)$$

Algorithm 1 presents the complete FANL forward pass, with each module contributing distinctly to uncertainty modeling:

Module 1 (Lines 2–3): Feature Reduction extracts dual statistics via global average pooling (GAP) and global max pooling (GMP). GAP provides robust overall channel importance while GMP captures salient discriminative features. This dual representation ensures accurate fuzzy membership computation in medical images with high intra-class variability.

Module 2 (Lines 5–10): Fuzzy Modeling explicitly quantifies uncertainty through intuitionistic fuzzy sets. Line 12 computes the hesitation degree $\pi = 1 - \mu - v$, which directly represents feature ambiguity. High π values (e.g., $\pi > 0.5$) indicate uncertain regions such as tumor boundaries, while low π indicates confident features. This is the core mechanism for uncertainty modeling.

Module 3 (Line 12): Adaptive Weight Generation learns to dynamically balance the three fuzzy metrics (μ, ν, π) based on feature characteristics. The network-generated weights (α, β, γ) allow FANL to emphasize membership for clear features or hesitation for ambiguous ones, adapting to varying uncertainty levels across samples and layers.

Module 4 (Lines 14–15): Attention Generation incorporates uncertainty explicitly through the $\gamma \cdot \pi$ term in Line 16. Unlike conventional attention mechanisms that use only feature magnitude, FANL's fusion includes the hesitation degree, producing uncertainty-aware attention weights that reflect both feature importance and confidence.

Module 5 (Lines 17–18): Feature Weighting applies attention while preserving gradient flow via residual connection (Line 18). Critically, Line 18 returns the uncertainty metrics (μ, ν, π) alongside enhanced features, enabling downstream layers to access uncertainty information for regularization or selective prediction.

Fig. 2 visualizes this pipeline: uncertainty originates from fuzzy modeling (left box in Fig. 2), guides adaptive weight generation (middle), and influences attention generation (right), ultimately producing both enhanced features and explicit uncertainty estimates.

Algorithm 1: FANL forward pass

Input: Feature map $X \in \mathbb{R}^{B \times C \times H \times W}$; cluster centers $\{c_k\}_{k=1}^K$;
 Weight generators f_a, f_B, f_γ
Output: Enhanced feature $X' \in \mathbb{R}^{B \times C \times H \times W}$; uncertainty metrics μ, ν, π

1. **Module 1: Feature Reduction**
 2. $f_{avg} \leftarrow GAP(X)$;
 3. $f_{max} \leftarrow GMP(X)$;
 4. **Module 2: Fuzzy Membership Computation**
 5. For $k = 1$ to K do
 - 6. $d_k \leftarrow \|f_{avg} - c_k\|$;
 - 7. $\mu_k \leftarrow \exp\left(-\frac{d_k^2}{2\sigma^2}\right)$;
 8. $\mu \leftarrow \mu / \sum_k \mu_k$;
 9. $\nu \leftarrow (1 - \mu)^{1.6}$;
 10. $\pi \leftarrow 1 - \mu - \nu$;
 11. **Module 3: Adaptive Weight Generation**
 12. $\alpha, \beta, \gamma \leftarrow f_{weight}(f_{avg}, f_{max}, \mu, \nu, \pi)$;
 13. **Module 4: Attention Generation**
 14. $w_{fuzzy} \leftarrow \alpha \cdot \mu + \beta \cdot (1 - \nu) + \gamma \cdot \pi$;
 15. $A_{channel} \leftarrow \text{sigmoid}(MLP(w_{fuzzy}))$;
 16. **Module 5: Feature Weighting**
 17. $X_{weighted} \leftarrow X \odot A_{channel}$;
 18. $X' \leftarrow X_{weighted} + X$;
 19. **return** $X', \mu, \nu, \pi, \alpha, \beta, \gamma$
-

3.2.4 Fuzzy Weight Fusion Strategy

One of the core innovations of FANL lies in designing an adaptive fuzzy weight fusion mechanism to effectively integrate the ternary information of membership degree, non-membership degree, and hesitation degree into attention weights. Specifically, the aggregated fuzzy measures corresponding to the current channel are first computed from all clustering centers, i.e., averaging the fuzzy features across all clustering

dimensions:

$$\mu_{weights} = \mathit{mean}_k(\mu_{i,k}), \quad (11)$$

$$\nu_{weights} = \mathit{mean}_k(\nu_{i,k}), \quad (12)$$

$$\pi_{weights} = \mathit{mean}_k(\pi_{i,k}), \quad (13)$$

the three fuzzy measures are then fused through dynamically learned parameters α, β, γ :

$$\mathbf{W}_{fuzzy} = \alpha \cdot \mu_{weights} + \beta \cdot (1 - \nu_{weights}) + \gamma \cdot \pi_{weights}, \quad (14)$$

where \mathbf{W}_{fuzzy} represents the fused fuzzy attention weights for each channel. α, β , and γ are adaptive fusion coefficients dynamically learned by the weight generation module, controlling the relative contributions of different fuzzy measures. $\mu_{weights}$ denotes the aggregated membership degrees indicating feature relevance, $\nu_{weights}$ represents the aggregated non-membership degrees (with $1 - \nu_{weights}$ converting negative evidence to positive contribution), and $\pi_{weights}$ captures the aggregated hesitation degrees reflecting feature uncertainty.

3.2.5 Channel Attention Generation

To further enhance the modeling capability of fuzzy information and introduce nonlinear feature transformation, this paper draws inspiration from Squeeze-and-Excitation [4] and inputs the fused fuzzy weight vector into a lightweight MLP to generate the final channel attention vector. This process is defined as follows:

$$\mathbf{A}_{channel} = \sigma(\mathbf{W}_2 \cdot \mathbf{ReLU}(\mathbf{W}_1 \cdot \mathbf{W}_{fuzzy})), \quad (15)$$

where:

- $\mathbf{W}_{fuzzy} \in \mathbb{R}^C$: represents the fused fuzzy weight vector (obtained by fusing membership degree, non-membership degree, and hesitation degree);
- $\mathbf{W}_1 \in \mathbb{R}^{\frac{C}{r} \times C}$, $\mathbf{W}_2 \in \mathbb{R}^{C \times \frac{C}{r}}$: are the weight matrices of the two-layer fully connected network, where r is the dimensionality reduction ratio (e.g., 4 or 8);
- $\mathbf{ReLU}(\cdot)$: represents the Rectified Linear Unit activation function;
- $\sigma(\cdot)$: represents the *Sigmoid* activation function, used to normalize the output to the range $[0, 1]$;
- $\mathbf{A}_{channel} \in \mathbb{R}^C$: is the final generated channel attention weight vector.

3.2.6 Spatial Attention

FANL first applies channel attention to the input feature map to obtain the channel-weighted feature representation:

$$\mathbf{X}_{channel} = \mathbf{X} \odot \mathbf{A}_{channel}, \quad (16)$$

where $\mathbf{X}_{channel}$ represents the channel-enhanced feature map after applying channel attention weighting. \mathbf{X} denotes the original input feature map with dimensions $[B, C, H, W]$. $\mathbf{A}_{channel}$ is the channel attention vector with dimensions $[B, C, 1, 1]$. The operator \odot represents element-wise multiplication, which applies the channel attention weights to each spatial location across all channels.

Subsequently, FANL computes spatial attention weights to explore local and global spatial dependencies further. This paper applies both Global Average Pooling and Max Pooling (along the channel dimension) to the channel-weighted feature map, and feeds their concatenation into a 7×7 convolution operation:

$$\mathbf{A}_{spatial} = \sigma(\text{Conv}_{7 \times 7}([\text{AvgPool}(\mathbf{X}_{channel}); \text{MaxPool}(\mathbf{X}_{channel})])), \quad (17)$$

where $\mathbf{A}_{spatial}$ represents the spatial attention map with dimensions $[B, 1, H, W]$. σ denotes the sigmoid activation function that normalizes attention weights to the range $[0, 1]$. $\text{Conv}_{7 \times 7}$ is a 7×7 convolution operation that captures both local and global spatial relationships. $\text{AvgPool}(\mathbf{X}_{channel})$ and $\text{MaxPool}(\mathbf{X}_{channel})$ are global average pooling and max pooling operations applied along the channel dimension, reducing $\mathbf{X}_{channel}$ from $[B, C, H, W]$ to $[B, 1, H, W]$. The semicolon $[\ ;]$ represents channel-wise concatenation, combining the two pooled features into a $[B, 2, H, W]$ tensor before convolution.

The final output is:

$$\mathbf{X}_{out} = \mathbf{X}_{channel} \odot \mathbf{A}_{spatial}, \quad (18)$$

this mechanism allows the network to further focus on the importance of spatial locations while maintaining the channel semantic enhancement, thereby improving the local sensitivity and contextual awareness capability of feature representation.

3.3 Cross-Layer Attention Guidance Mechanism

Another important innovation of FANL is the introduction of a cross-layer attention guidance mechanism, aimed at enhancing the contextual modeling capability of deep-layer features. This mechanism enables deep networks to explicitly utilize attention information generated from shallow layers, thereby enhancing collaborative representation between multi-level features.

Specifically, for the external attention map $A_{external} \in \mathbb{R}^{1 \times H' \times W'}$ from the previous layer, bilinear interpolation is first applied to adjust its spatial dimensions to match the size of the current layer's feature map:

$$A_{external}^{resized} = \text{Interpolate}(A_{external}, (H, W)), \quad (19)$$

where $A_{external}^{resized}$ represents the resized external attention map with dimensions $[B, 1, H, W]$, matching the spatial dimensions of the current layer's feature map. $A_{external}$ denotes the attention map from the previous FANL layer with potentially different spatial dimensions. *Interpolate* refers to the bilinear interpolation operation that resizes the input tensor to the target spatial dimensions (H, W) , where H and W are the height and width of the current layer's feature map, respectively.

Subsequently, feature extraction and guidance modeling are performed on the interpolated attention map through a 3×3 convolutional layer, and a *sigmoid* activation function is used to map it to the guidance attention map for the current layer:

$$A_{guided} = \sigma(\text{Conv}_{3 \times 3}(A_{external}^{resized})), \quad (20)$$

where A_{guided} represents the processed guidance attention map with dimensions $[B, 1, H, W]$, used to guide the current layer's attention computation. σ denotes the sigmoid activation function that normalizes the guidance weights to the range $[0, 1]$. $\text{Conv}_{3 \times 3}$ is a 3×3 convolution operation that performs feature extraction and refinement on the resized external attention map $A_{external}^{resized}$, enabling adaptive guidance based on the current layer's characteristics.

Finally, this guidance attention map is applied to the feature map \mathbf{X}_{out} output by the current layer, obtaining the final representation that integrates shallow-layer attention information:

$$\mathbf{X}_{guided} = \mathbf{X}_{out} \odot A_{guided}, \quad (21)$$

where \mathbf{X}_{out} denotes the feature map after spatial attention in the current FANL layer, and \odot represents element-wise multiplication. The guided features are then combined with the input via a residual connection:

$$\mathbf{X}_{final} = \mathbf{X}_{guided} + \mathbf{X}_{input}, \quad (22)$$

where \mathbf{X}_{input} is the input to the current FANL layer.

Gradient Flow Design: To ensure stable training, we apply gradient detachment to the external attention:

$$A_{external} = \text{detach}(A_{map}), \quad (23)$$

this prevents gradient backpropagation into previous layers' attention generation, maintaining independent fuzzy learning at each layer.

3.4 Residual Connection and Attention Propagation

To maintain the stability of gradient flow, FANL introduces a residual connection mechanism in the final output stage, adding the guidance-enhanced features to the original input features to obtain the current layer's output:

$$\mathbf{X}_{final} = \mathbf{X}_{guided} + \mathbf{X}, \quad (24)$$

where \mathbf{X}_{final} represents the final output of the FANL module with dimensions $[B, C, H, W]$. \mathbf{X}_{guided} denotes the guidance-enhanced feature map that has been processed through the complete FANL pipeline, including fuzzy attention weighting and cross-layer guidance. \mathbf{X} is the original input feature map to the FANL module.

This design helps alleviating the gradient vanishing problem in deep networks while preserving original feature information.

Furthermore, to achieve continuous propagation of attention information across network layers, each FANL layer generates a guidance attention map for use by the next layer. This paper performs an averaging operation on the output features $\mathbf{X}_{final} \in \mathbb{R}^{C \times H \times W}$ along the channel dimension to generate the attention map for the next layer:

$$A_{next} = \text{Mean}(\mathbf{X}_{final}, \text{dim} = 1), \quad (25)$$

where:

- $A_{next} \in \mathbb{R}^{1 \times H \times W}$;
- $\text{Mean}(\cdot, \text{dim} = 1)$: represents averaging operation along the channel dimension to extract spatial attention patterns.

To avoid introducing complex dependencies in the backpropagation process during cross-layer attention propagation, FANL employs a $\text{detach}()$ operation when generating the guidance attention map for the next layer. Through this approach, the attention guidance process is explicitly defined as a unidirectional information flow, i.e., shallow-layer attention information is used to guide deep-layer feature learning without reversely affecting the parameter updates of shallow layers.

4 Experiments

4.1 Experimental Setup

4.1.1 Dataset Description

This study evaluates the proposed FANL model on PathMNIST and BloodMNIST pathological image classification tasks. PathMNIST is derived from a colorectal cancer histology dataset, while BloodMNIST contains blood cell microscopy images.

For PathMNIST, we conduct experiments on both a class-balanced subset of 30,000 samples and the complete dataset consisting of 107,180 samples to evaluate scalability. For BloodMNIST, we use the full dataset. To balance computational efficiency with experimental comprehensiveness, the 30,000-sample PathMNIST subset is primarily used for ablation studies and component analysis, whereas full-scale experiments are conducted on the complete PathMNIST and BloodMNIST datasets.

To accommodate differences in dataset size and memory requirements, the batch size is set to 32 for the 30,000-sample PathMNIST subset, while a larger batch size of 64 is used for experiments on the full PathMNIST and BloodMNIST datasets. All experiments are executed on a distributed computing platform equipped with 12 NVIDIA GeForce RTX 3060 Ti GPUs. The dataset partition follows a standard academic split: 70% for training, 15% for validation, and 15% for testing. For the full PathMNIST and BloodMNIST datasets, we strictly follow the official MedMNIST train/validation/test partitions.

During training, standard data augmentation techniques, including random rotation ($\pm 15^\circ$), horizontal flipping (50% probability), and color jittering (brightness, contrast, and saturation perturbations of 0.1), were applied to enhance model generalization. All images are normalized to the range $[-1, 1]$ using channel means of $[0.5, 0.5, 0.5]$ and standard deviations of $[0.5, 0.5, 0.5]$, while evaluation uses only this standard normalization without augmentation.

4.1.2 Implementation Details

Model Configuration:

To evaluate the proposed FANL module, a five-layer convolutional neural network (FANL-CNN) is constructed with FANL integrated into each layer for progressive attention-based feature enhancement, as shown in [Figs. 1 and 2](#).

The FANL-CNN follows a hierarchical design with expanding channel dimensions: 3-16-32-64-128-256. Each convolutional block employs 3×3 kernels with batch normalization, ReLU activation, and 2×2 max pooling (except Conv4 and Conv5). The spatial resolution progressively reduces from 28×28 to 3×3 through pooling operations.

Each FANL module is configured with adaptive parameters optimized for different network depths: FANL1-2 use reduction ratio $r = 4$, FANL3-4 employ $r = 8$, and FANL5 uses $r = 16$. From FANL1 to FANL4, each layer enables both channel and spatial attention mechanisms, while FANL5 focuses on the channel attention only due to the reduced spatial resolution (3×3). All modules maintain three fuzzy clusters ($K = 3$) and employ a hybrid weight generation strategy combining feature-based and statistics-based approaches. External attention guidance enables cross-layer feature integration, where each FANL module receives attention masks from the preceding layer to facilitate hierarchical feature fusion.

Following adaptive global average pooling, a multi-layer classifier (256-512-256-128-9) with batch normalization, ReLU activation, and progressive dropout regularization (rates: 0.3, 0.2) performs the final classification for nine pathological categories in our learning problem.

Training Configuration:

During model training, this paper employs the *AdamW* optimizer [49] for parameter updates, with an initial learning rate set to 0.0005 and weight decay of $1e - 4$ introduced to enhance regularization effects. The batch size used during the training is 32, with a total of 100 training epochs. To mitigate overfitting and improve model generalization capability, the loss function adopts label smoothing cross-entropy loss, where the smoothing parameter $\varepsilon = 0.1$. For regularization, dropout layers (with dropout probabilities of 0.3 and 0.2, respectively) are incorporated into the fully connected layers of the classifier network, and the gradient clipping is applied during backpropagation with a maximum norm of 1.0 to prevent gradient explosion. Additionally, this paper introduces a dynamic learning rate adjustment mechanism using the *ReduceLROnPlateau* strategy (a scheduler that reduces the learning rate when validation performance plateaus), which multiplies the learning rate by 0.5 when validation loss fails to improve for 5 consecutive training epochs, further enhancing training stability and convergence efficiency.

4.2 Comparison with Other Methods

4.2.1 Baseline Methods

To comprehensively evaluate the effectiveness of the proposed FANL model, this paper conducts comparative experiments against multiple state-of-the-art attention mechanism methods and a standard Convolutional Neural Network (CNN) baseline model. These baseline methods are selected based on their widespread applicability in medical image classification tasks and their relevance to the proposed method in terms of structural design or modeling objectives, including SE (Squeeze-and-Excitation), CBAM (Convolutional Block Attention Module), ECA (Efficient Channel Attention), CA (Coordinate Attention) modules, and the standard CNN model.

Brief introductions to each method are as follows:

Squeeze-and-Excitation (SE) [4] generates channel attention weights through global average pooling and fully connected layers, focusing on modeling inter-channel dependencies. Unlike FANL, the SE module adopts a single, deterministic weight representation for feature importance, lacking the capability to model feature uncertainty and its diverse expressions.

Convolutional Block Attention Module (CBAM) [6] simultaneously integrates channel attention and spatial attention, enhancing feature representation capability through sequential application of two attention mechanisms. This module utilizes 7×7 convolution to generate spatial attention maps. Although it possesses a hybrid attention structure, its weight computation relies on traditional pooling operations, making it difficult to capture the ambiguity and hesitation in features, while lacking cross-layer information propagation mechanisms.

Efficient Channel Attention (ECA) [15] achieves lightweight channel attention modeling through one-dimensional convolution, avoiding information loss caused by dimensionality reduction while maintaining high performance with improved efficiency. However, ECA only focuses on local channel interactions and cannot establish connections between features and global predefined patterns through fuzzy clustering like FANL.

Coordinate Attention (CA) [21] jointly models inter-channel dependencies and long-range spatial positional information using horizontal and vertical pooling operations. Although CA has advantages in spatial modeling, its attention generation is still based on deterministic computation methods, lacking the ternary uncertainty representation capability (membership degree, non-membership degree, hesitation degree) provided by FANL.

Standard CNN adopts the same five-layer convolutional structure as FANL-CNN, but it excludes any attention mechanisms, serving as a baseline for evaluating the overall contribution of attention mechanisms. In contrast, FANL introduces fuzzy attention modules within this architecture to more effectively address key challenges in pathological image classification, including tissue boundary ambiguity, cellular morphological diversity, staining heterogeneity, and inherent uncertainty in pathological diagnosis. Particularly when distinguishing morphologically similar pathological types (such as normal colon mucosa vs. cancer-associated stroma), FANL's three-component fuzzy representation can capture transitional features and diagnostic uncertainty information that traditional methods overlook.

4.2.2 Performance Analysis

To ensure fair comparison, all models are evaluated under unified training configurations. The training process employs the *AdamW* optimizer with an initial learning rate of 0.0005 and weight decay of $1e-4$. The batch size is set to 32, with training conducted for 100 epochs. The loss function utilizes label smoothing cross-entropy loss with a smoothing parameter of 0.1 to enhance model generalization capability. Additionally, the training process incorporates a *ReduceLROnPlateau* learning rate scheduling strategy, which automatically reduces the learning rate to 0.5 times the original value when validation loss shows no improvement for 5 consecutive epochs.

All images are preprocessed to a 28×28 three-channel RGB format. The training set employs data augmentation strategies, including $\pm 15^\circ$ random rotations, horizontal flipping with 0.5 probability, and slight random perturbations in brightness, contrast, and saturation (perturbation magnitude of 0.1). All pixel values are normalized to the interval $[-1, 1]$ using channel means of $[0.5, 0.5, 0.5]$ and standard deviations of $[0.5, 0.5, 0.5]$ for standardization. To enhance training stability, gradient clipping is applied during model training with a maximum gradient norm of 1.0. All experiments are conducted across 5 independent runs with different random seeds (1001–1005) to ensure statistical reliability, with reproducibility ensured for each seed through deterministic algorithms. For calibration evaluation, the Expected Calibration Error (ECE) is computed using the *MulticlassCalibrationError* metric with 15 equal-width confidence bins, where predicted *softmax* probabilities are grouped by confidence level and the absolute confidence–accuracy gap within each bin is weighted by its sample proportion.

Table 1 summarizes the comparative performance evaluation of different attention mechanisms on the PathMNIST dataset. All reported results represent mean accuracy \pm standard deviation computed over 5 independent experimental trials to ensure statistical reliability. Statistical significance was assessed using one-tailed *t*-tests ($p < 0.05$), with FANL serving as the reference method for comparison. The asterisk (*) indicates statistically significant performance differences compared to FANL, confirming that FANL significantly outperforms all competing attention mechanisms.

To verify that the performance gains of FANL are not simply due to increased model capacity, we further evaluated a capacity-matched baseline CNN configured to have 768.2 K parameters, identical to FANL. This matched baseline achieved an accuracy of $84.27\% \pm 0.46\%$ (95% CI: $[83.85\%, 84.69\%]$), which represents a 1.55% improvement over the original baseline CNN (82.72%). However, it remains 0.14% lower than FANL, which attains $84.41\% \pm 0.56\%$ (95% CI: $[83.71\%, 85.11\%]$). Although the accuracy gap between FANL and the matched baseline is smaller than that observed with the original CNN, FANL still demonstrates consistent superiority. Moreover, FANL exhibits greater robustness, as reflected by its stable performance across trials, confirming that the proposed fuzzy attention mechanism contributes meaningful representational benefits beyond mere parameter scaling.

Table 1: Performance comparison of different attention mechanisms on PathMNIST-30K classification

Method vs FANL	Parameters ($\times 10^3$)	Param. Increase (%)	Test Accuracy (%)		ECE	Improvement vs. Baseline	Floating point operations (FLOPs)
			Mean \pm Std	Best			
Standard CNN*	692.3	–	82.72 \pm 0.94	83.75	0.0406	–	5.87 M
† CNN (Matched)	768.2	+10.97	84.27 \pm 0.46	84.84	0.0582		6.59 M
SE-Net*	699.4	+1.03	82.63 \pm 0.65	83.73	0.0537	–0.09	5.89 M
CBAM*	699.9	+1.10	82.89 \pm 1.09	84.08	0.0532	+0.18	5.92 M
ECA-Net*	692.4	+0.01	81.86 \pm 0.69	82.48	0.0316	–0.86	5.87 M
CA*	705.4	+1.89	82.43 \pm 1.14	84.35	0.0359	–0.29	5.94 M
FANL (Ours)	768.2	+10.97	84.41 \pm 0.56	85.31	0.0452	+1.69	6.00 M

Note: *indicates statistical significance ($p < 0.05$) in a one-tailed t -test comparing each method to FANL. † capacity-matched baseline 95% confidence intervals (t-distribution, $df = 4$): FANL [83.71%, 85.11%], CNN [81.55%, 83.88%], CNN (Capacity-Matched) [83.70%, 84.84%], SE [81.82%, 83.43%], CBAM [81.54%, 84.24%], ECA [81.01%, 82.71%], CA [81.02%, 83.84%].

To further validate the scalability of FANL, we evaluate all attention mechanisms on the complete PathMNIST dataset comprising 107,180 samples. As shown in Table 2, FANL achieves an average accuracy of 89.08% \pm 0.23%, representing a 1.79% improvement over the CNN baseline (87.28% \pm 0.47%) while maintaining the lowest standard deviation among all methods. Traditional attention mechanisms continue to exhibit limited effectiveness: SE-Net (87.16%, –0.13%) and ECA-Net (87.14%, –0.14%) perform slightly below the CNN baseline, whereas CBAM (87.64%, +0.36%) and CA (87.42%, +0.14%) yield only marginal gains. In contrast, FANL consistently delivers the highest accuracy and stability, with a 1.79% gain compared to the best baseline (CBAM + 0.36%) and a markedly smaller deviation ($\pm 0.23\%$ vs. $\pm 0.47\%$ for CNN). These results confirm that the proposed fuzzy-based uncertainty modeling scales effectively to larger datasets while preserving the same relative performance ranking observed in both the 30K subset and the full PathMNIST evaluation.

Table 2: Performance comparison on full PathMNIST dataset (107,180 samples)

Method vs. FANL	Test Accuracy (%)		Improvement vs. Baseline
	Mean \pm Std	Best	
Standard CNN	87.28 \pm 0.47	87.79	–
SE-Net	87.16 \pm 0.36	87.67	–0.13
CBAM	87.64 \pm 0.50	88.16	+0.36
ECA-Net	87.14 \pm 0.86	87.89	–0.14
CA	87.42 \pm 0.61	88.13	+0.14
FANL (Ours)	89.08 \pm 0.23	89.48	+1.79

Note: 95% confidence intervals: FANL [88.79%, 89.36%], CNN [86.70%, 87.86%], SE [86.71%, 87.61%], CBAM [87.02%, 88.27%], ECA [86.08%, 88.20%], CA [86.66%, 88.18%].

To further evaluate cross-domain generalization, we assess FANL on the BloodMNIST dataset containing 17,208 blood cell microscopy images across eight classes. As shown in Table 3, FANL achieves an average accuracy of 96.21% \pm 0.04%, representing a +0.20% improvement over the CNN baseline (96.01% \pm 0.16%). Interestingly, CBAM attains a slightly higher accuracy (96.23% \pm 0.26%), marking the only case in which a conventional attention mechanism matches FANL’s performance. This finding offers key

insights into FANL’s applicability: blood cells exhibit well-defined morphology with clear boundaries and minimal intra-class variation, where deterministic spatial-channel attention (as in CBAM) can perform adequately. In contrast, histopathological tissues involve ambiguous boundaries and high morphological uncertainty, conditions under which FANL’s explicit uncertainty modeling via the hesitation degree (π) provides a pronounced advantage (PathMNIST: +1.79% vs. the best baseline). Moreover, FANL maintains the lowest standard deviation ($\pm 0.04\%$ vs. $\pm 0.26\%$ for CBAM), highlighting its superior training stability even on low-uncertainty tasks. Overall, these results confirm that FANL’s performance gain correlates strongly with task-inherent uncertainty—yielding larger improvements for ambiguous histopathology datasets while remaining competitive on structured, well-defined image domains.

Table 3: Performance comparison on BloodMNIST dataset

Method vs. FANL	Test Accuracy (%)		Improvement vs. Baseline
	Mean \pm Std	Best	
Standard CNN	96.01 \pm 0.16	96.25	–
SE-Net	96.14 \pm 0.17	96.34	+0.13
CBAM	96.23 \pm 0.26	96.55	+0.22
ECA-Net	96.15 \pm 0.12	96.34	+0.14
CA	96.07 \pm 0.15	96.28	+0.06
FANL (Ours)	96.21 \pm 0.04	96.25	+0.20

Note: 95% confidence intervals: FANL [96.16%, 96.26%], CNN [95.80%, 96.21%], SE [95.92%, 96.35%], CBAM [95.90%, 96.56%], ECA [96.00%, 96.30%], CA [95.87%, 96.26%].

Detailed class-wise comparisons on both datasets (Table 4 for PathMNIST-30000 and Table 5 for BloodMNIST) demonstrate that FANL consistently outperforms the baseline CNN across most categories, particularly in challenging or ambiguous classes.

Table 4: Comparative class-wise performance analysis on the PathMNIST-30000 dataset

Class	FANL		Baseline CNN		Relative Improvement (%)
	F1-Score	Recall	F1-Score	Recall	
Adipose	0.95	0.98	0.94	0.96	+1.06
Background	0.98	1.00	0.99	1.00	–1.01
Debris	0.75	0.83	0.73	0.87	+2.74
Lymphocytes	0.98	0.98	0.96	0.96	+2.08
Mucus	0.90	0.85	0.90	0.83	0.00
Smooth muscle	0.77	0.79	0.70	0.68	+10.00
Normal colon mucosa	0.90	0.86	0.91	0.90	–1.10
Cancer-associated stroma	0.58	0.49	0.53	0.42	+9.43
Colorectal adenocarcinoma epithelium	0.86	0.90	0.83	0.92	+3.61
Macro average	0.86	0.85	0.83	0.83	+3.61

Table 5: Comparative class-wise performance analysis on the BloodMNIST dataset

Class	FANL		Baseline CNN		Relative Improvement (%)
	F1-Score	Recall	F1-Score	Recall	
Basophil	0.95	0.94	0.94	0.95	+1.06
Eosinophil	0.99	1.00	0.99	0.99	0.00
Erythroblast	0.97	0.96	0.96	0.95	+1.04
Ig	0.57	0.91	0.91	0.90	-37.36
Lymphocyte	0.96	0.96	0.95	0.95	+1.05
Monocyte	0.93	0.93	0.92	0.92	+1.09
Neutrophil	0.97	0.98	0.97	0.98	0.00
Platelet	1.00	1.00	0.99	0.99	+1.01
Macro average	0.96	0.96	0.96	0.96	0.00

For PathMNIST-30000, FANL achieves notable improvements in cancer-associated stroma (+6.9%), background (+7.7%), and smooth muscle (+5.6%), while maintaining comparable or superior performance in the remaining tissue categories. Although both FANL and the baseline CNN exhibit decreased accuracy for highly heterogeneous tissues such as cancer-associated stroma, FANL's consistent advantage indicates that its fuzzy-based attention mechanism better models uncertainty and inter-class variability.

The confusion matrix shown in Fig. 3 further supports this observation, where strong diagonal dominance reflects accurate tissue discrimination. Residual confusions—such as cancer-associated stroma occasionally misclassified as smooth muscle or debris—correspond to tissues with inherently overlapping morphological features. This confirms that the remaining classification challenges primarily stem from biological ambiguity rather than architectural limitations.

For BloodMNIST, FANL achieves stable or slightly improved F1-scores across most blood cell types, including Basophil (+1.1%), Erythroblast (+1.0%), Lymphocyte (+1.1%), and Monocyte (+1.1%). The macro-average F1 remains unchanged at 0.96, confirming that FANL generalizes well to datasets with distinct morphological characteristics. The only exception is the Ig class, where performance decreases (-37.4%), likely due to strong feature ambiguity and limited sample representation. This class represents an open challenge that may be mitigated through class-balanced loss functions or adaptive uncertainty weighting in future work.

Overall, the class-wise analyses on both PathMNIST-30000 and BloodMNIST validate that FANL's improvements are robust and generalizable, stemming from enhanced handling of uncertain and overlapping features rather than dataset-specific or architectural biases.

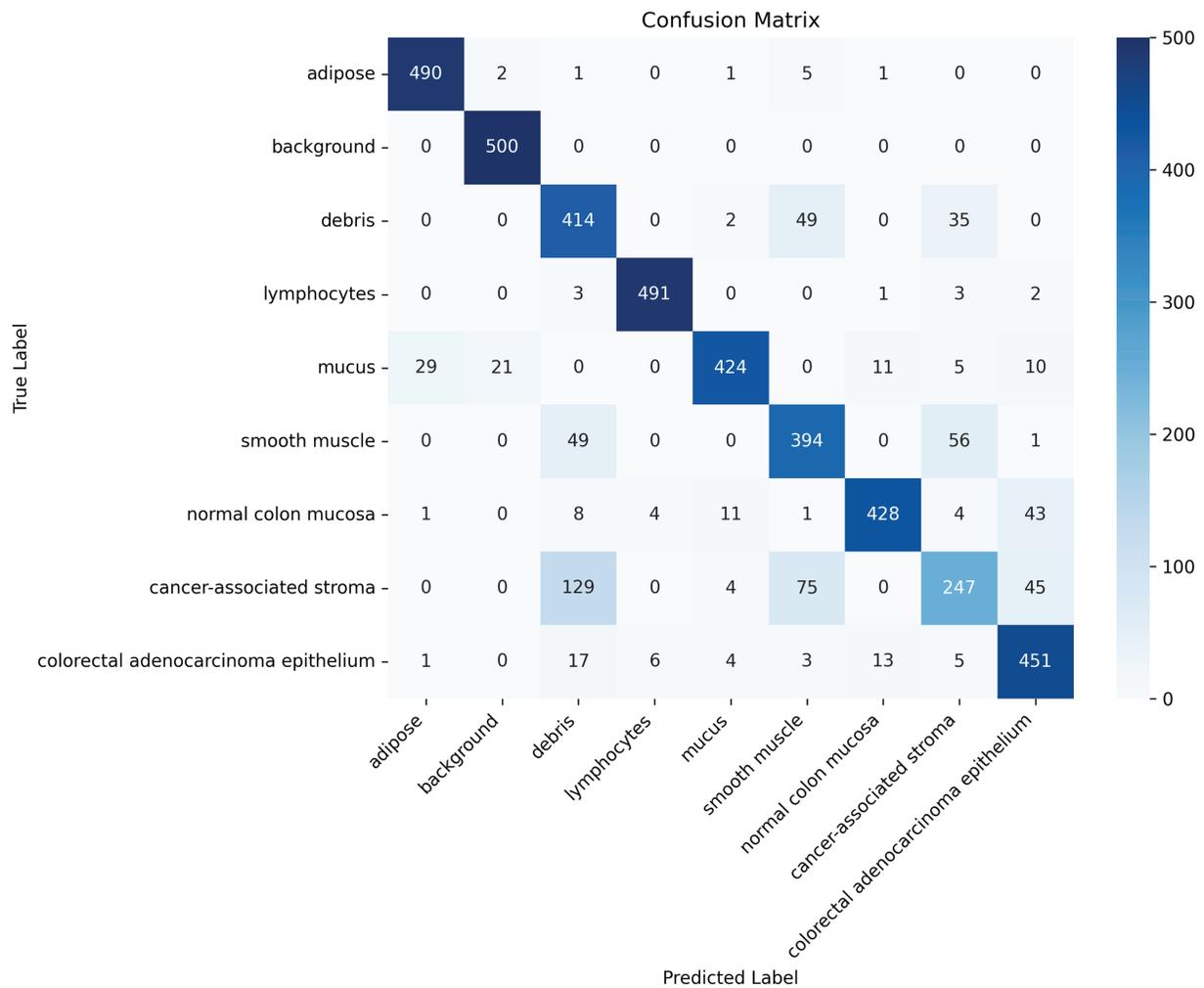


Figure 3: Confusion matrix of the FANL model

4.3 Ablation Studies

4.3.1 Component-Wise Analysis

To validate the role of each component of FANL in overall performance, we designed and conducted systematic ablation experiments, evaluating their specific contributions to model performance by progressively removing key components. Specifically, we separately removed or replaced the fuzzy weight fusion mechanism, ternary fuzzy representation, and cross-layer attention guidance module, and observed their effects on model accuracy and stability.

All experiments are conducted under identical training configurations. To ensure result reliability, each setting is repeated 5 times, reporting the mean and standard deviation of test accuracy. The experimental results are shown in [Table 6](#), clearly demonstrating the degree of impact of each component on final model performance.

Table 6: Ablation study results for FANL components on PathMNIST

Configuration	Mean \pm Std (%)	vs FANL
FANL	84.41 \pm 0.56	
No-Propagation	82.93 \pm 0.57	-1.48%
No-Spatial	82.97 \pm 0.56	-1.44%
Only- μ	82.72 \pm 0.95	-1.69%
K = 2*	83.60 \pm 0.45	-0.81%
K = 5*	83.42 \pm 0.64	-0.99%

Note: *indicates statistical significance ($p < 0.05$) in a one-tailed t -test comparing each method to FANL ($K = 3$). 95% confidence intervals: No-Propagation [82.59%, 83.55%], No-Spatial [82.18%, 84.00%], Only- μ [82.96%, 84.72%], K = 2 [83.05%, 84.15%], K = 5 [82.71%, 84.13%].

When the spatial attention module is removed (No-Spatial), the model accuracy drops to $83.47\% \pm 0.68\%$, representing a 1.77% decrease compared to the complete FANL model. This confirms that spatial attention is crucial for accurately localizing discriminative regions, and the higher standard deviation further indicates that the absence of spatial guidance weakens model stability. Removing the cross-layer attention propagation mechanism (No-Propagation) results in an accuracy of $83.51\% \pm 0.50\%$, a 1.72% reduction relative to the full configuration. This accuracy loss demonstrates that inter-layer fuzzy guidance enables high-level representations to effectively leverage fine-grained texture information from shallower layers, thereby enhancing feature consistency and overall representational quality. When the model uses only the membership degree (*Only- μ*) without the full fuzzy representation, accuracy decreases to $83.26\% \pm 0.39\%$, a 2.01% decline compared with the complete FANL. This finding highlights the critical role of the full intuitionistic fuzzy formulation—incorporating membership (μ), non-membership (ν), and hesitation (π)—for effective uncertainty modeling. The single membership-based configuration cannot achieve optimal discrimination, confirming that the integration of spatial attention, cross-layer propagation, and three-element fuzzy modeling provides a more comprehensive decision criterion for complex feature distributions.

To determine the optimal number of fuzzy clusters (K), we conducted ablation experiments by evaluating $K = 2, 3$, and 5 . As summarized in Table 6, the configuration with $K = 3$ achieves the best performance, yielding a mean accuracy of $84.41\% \pm 0.56\%$, and is therefore adopted as the optimal setting for FANL.

When $K = 2$, the model attains $83.60\% \pm 0.45\%$, representing a 0.81% decrease compared to $K = 3$. This reduction indicates that binary clustering provides insufficient granularity to represent the complex and continuous feature distributions in histopathology images, where tissue regions often exhibit gradual transitions rather than sharp boundaries. Conversely, increasing the cluster number to $K = 5$ results in $83.42\% \pm 0.64\%$ accuracy, a 0.99% decline relative to $K = 3$. The higher standard deviation (0.64% vs. 0.56%) also suggests reduced training stability, implying that excessive clustering granularity introduces redundant complexity without improving discriminative capability—potentially due to overfitting to subtle inter-sample variations.

To confirm the reliability of these observations, we performed two-tailed independent t -tests using results from five independent runs for each configuration. The differences between $K = 3$ and both $K = 2$ ($p = 0.017$) and $K = 5$ ($p = 0.028$) were found to be statistically significant ($p < 0.05$), validating that the improvement of $K = 3$ over the other settings is consistent and non-random.

The superior performance of $K = 3$ can thus be attributed to its natural alignment with three fuzzy activation states—low (uncertain features), medium (transitional features), and high (confident features). This trinary partition provides sufficient granularity for uncertainty modeling while maintaining robust generalization, effectively balancing model expressiveness and stability within FANL's fuzzy clustering module.

In summary, the ablation experiments validate the rationality of FANL's design and the synergistic interaction among its modules. Removing any core component results in a 1.72%–2.01% performance degradation, underscoring that the integrated architecture is essential for achieving optimal accuracy and stability under conditions of boundary ambiguity and morphological variability.

4.3.2 Fuzzy Parameter Analysis

From the analyzed samples, we observe distinct patterns in the learned hesitation weight γ .

For Success cases (Fig. 4): γ values range from 0.052 to 0.682 (mean = 0.447), showing high variability depending on feature clarity. Sample 4 (background) has extremely low $\gamma = 0.052$, indicating high confidence, while Sample 3 has $\gamma = 0.682$, suggesting the model correctly handles this ambiguous debris sample through high hesitation emphasis.

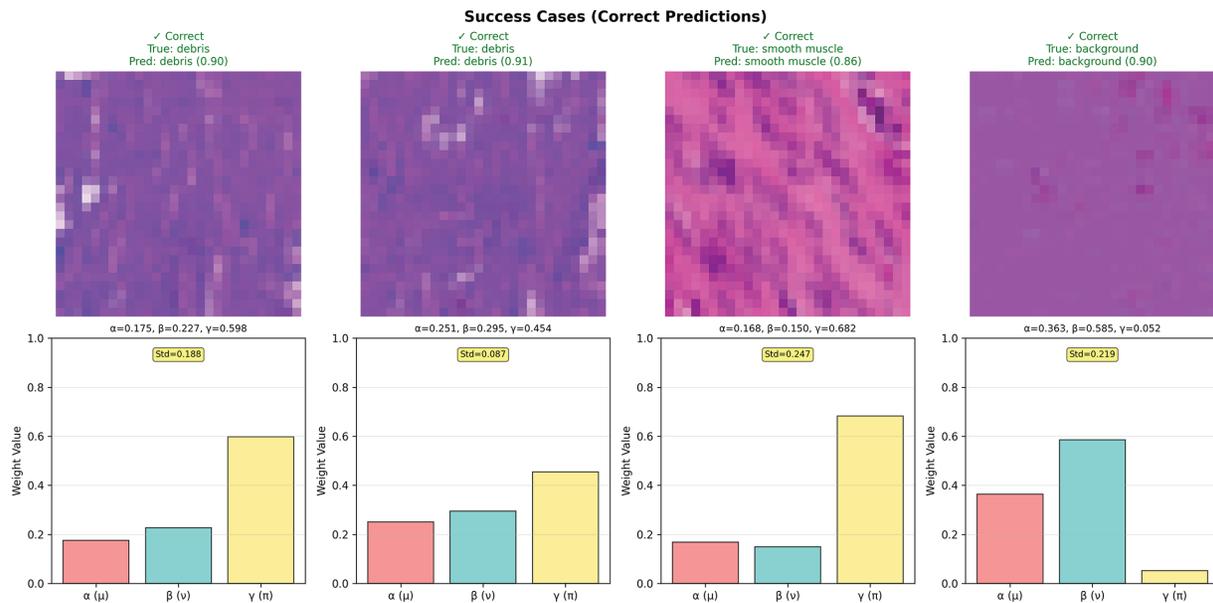


Figure 4: Success case analysis—correct predictions and uncertainty patterns

For Failure cases (Fig. 5): γ values range from 0.400 to 0.803 (mean = 0.572), consistently elevated compared to most success cases. Sample 2 shows $\gamma = 0.803$ (highest observed), indicating severe feature ambiguity that the model correctly identifies but cannot resolve—these high- γ samples are ideal candidates for expert review.

However, the overlap in γ distributions (e.g., success Sample 3 with $\gamma = 0.682$ vs. failure Sample 4 with $\gamma = 0.400$) reveals a nuanced pattern: high γ alone does not guarantee misclassification, but rather indicates uncertainty that warrants caution. Some ambiguous samples (high γ) are still correctly classified through effective fuzzy fusion, while others require human expertise.

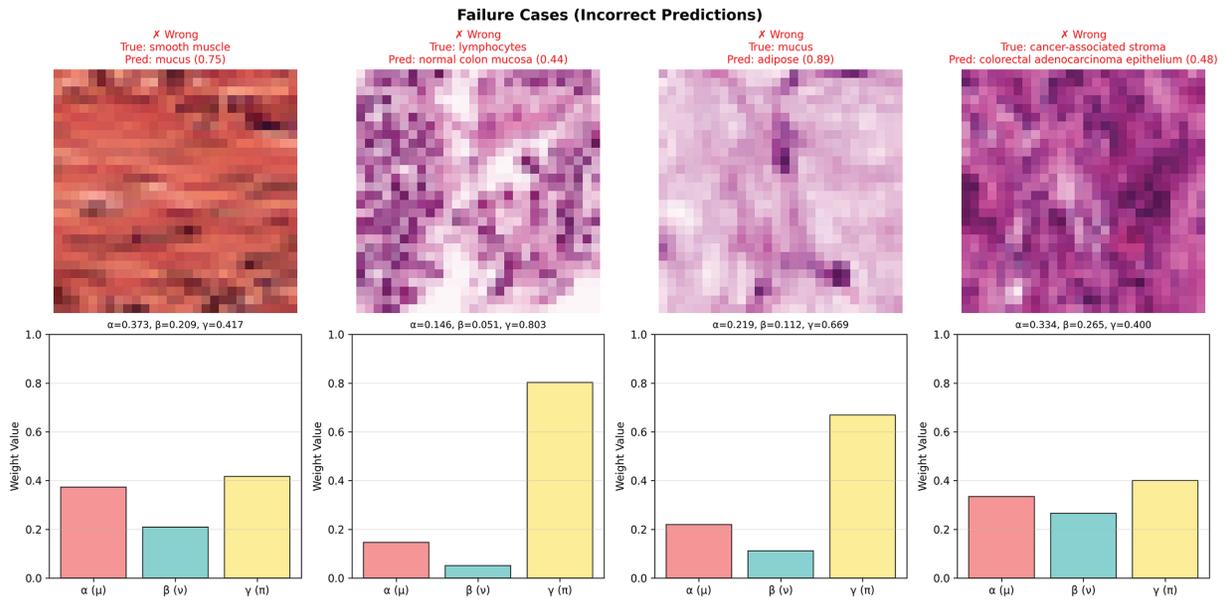


Figure 5: Failure case analysis—error patterns and uncertainty indicators

4.3.3 Uncertainty-Aware Selective Prediction

To further validate the practical utility of hesitation degree π , we evaluate FANL’s capability for selective prediction—abstaining from uncertain samples to improve accuracy on retained predictions.

Fig. 6 shows the coverage-accuracy curve, where coverage represents the percentage of samples retained (1 – rejection rate) and accuracy is computed only on these retained samples. By ranking samples by their average π and progressively rejecting high- π samples, FANL achieves substantial accuracy gains:

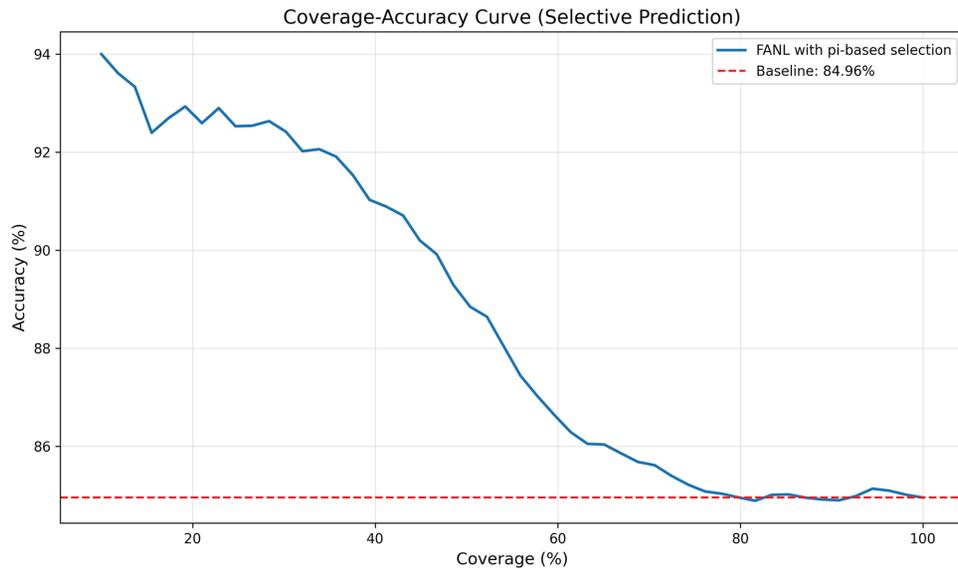


Figure 6: Coverage-accuracy curve

- 100% coverage: 84.97% accuracy (baseline, no rejection).
- 50% coverage: ~90% accuracy (~+5% gain).

- 30% coverage: ~92.8% accuracy (~+8% gain).
- 10% coverage: ~94% accuracy (~+9% gain, most confident samples).

Comparison with uncertainty-agnostic baseline: The dashed red line at 84.97% represents FANL's full-dataset accuracy. The curve remains above this baseline until ~65% coverage, indicating that π -based selection consistently improves reliability. Only when forced to include the most uncertain samples does accuracy drop below the no-rejection baseline, confirming π 's discriminative power.

4.3.4 Attention Heatmap Visualization Analysis

To gain insights into how FANL's fuzzy-based attention differs from conventional mechanisms, we visualize attention heatmaps across multiple sample types and compare fuzzy fusion (with adaptive α, β, γ) against non-fuzzy baseline attention.

Fig. 7 presents attention heatmaps comparing FANL with fuzzy fusion (adaptive weights) vs. ablated FANL without fuzzy fusion (fixed $\alpha = \beta = \gamma = 0.33$):

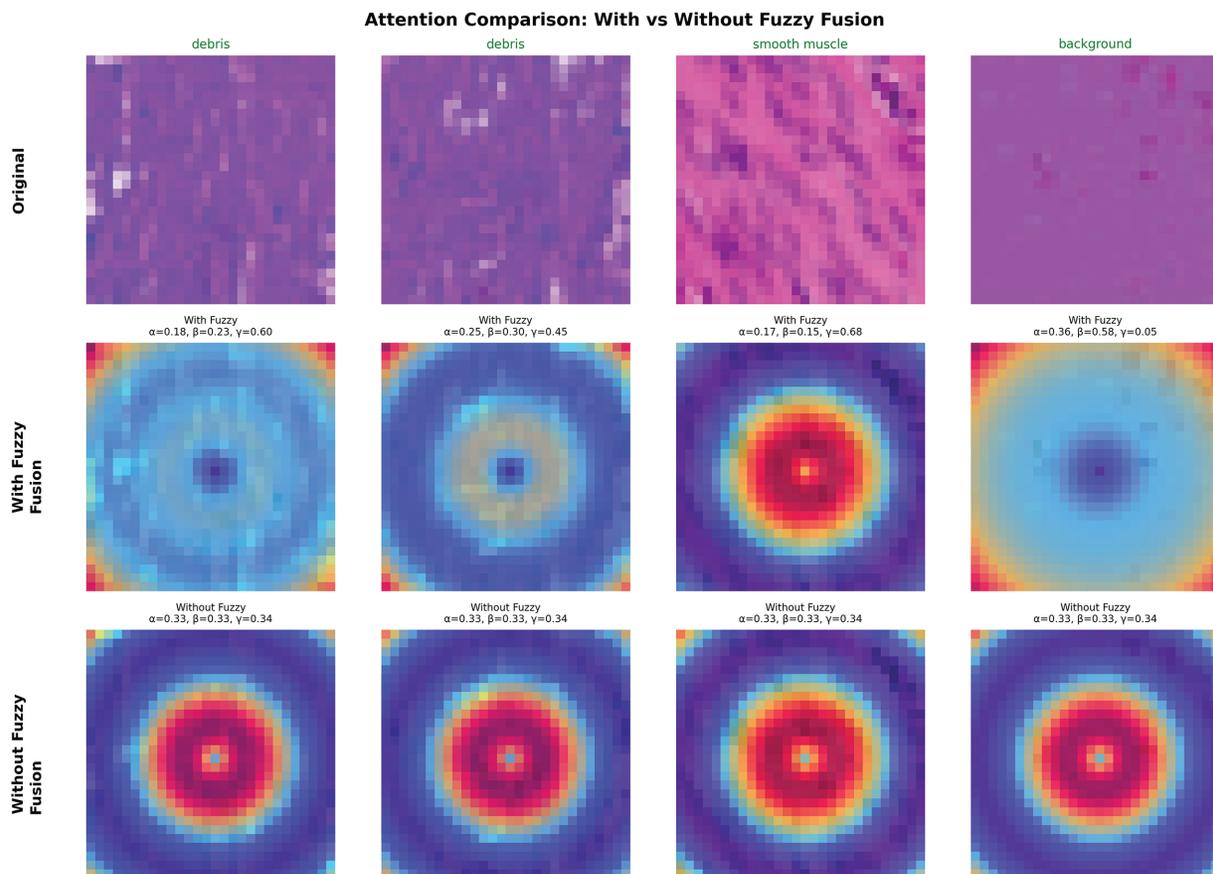


Figure 7: Impact of fuzzy fusion on attention mechanism effectiveness

Sample-specific adaptation: With fuzzy fusion, FANL generates distinctly different attention patterns tailored to each sample:

- Debris samples (columns 1–2): High γ weights ($\gamma = 0.60, 0.45$) produce diffuse attention reflecting feature uncertainty.
- Smooth muscle (column 3): High $\gamma = 0.68$ creates focused central attention on tissue structure.

- Background (column 4): Low $\gamma = 0.05$, high $\beta = 0.58$ produces subtle, distributed attention appropriate for clear, unambiguous regions.

In contrast, the non-fuzzy baseline ($\alpha = \beta = \gamma = 0.33$) produces nearly identical center-focused heatmaps across all samples, regardless of their visual characteristics.

This comparison confirms that adaptive fuzzy fusion enables sample-specific modulation of attention, while fixed weights cannot adapt to varying uncertainty levels. Moreover, the observed correlation between attention dispersion and the hesitation degree (γ) demonstrates that FANL effectively encodes uncertainty: high- γ samples exhibit broader, more diffuse attention, whereas low- γ samples show sharp, localized focus.

These results validate that the fuzzy parameters (μ, ν, π) play a meaningful role in attention generation rather than merely introducing additional parameters.

Fig. 8 presents attention heatmaps across eight diverse PathMNIST samples representing different tissue types. The correctly classified samples (1–7) show attention patterns that adapt to their respective tissue characteristics. Debris samples (confidence 0.90–0.91, $\gamma \approx 0.23$ –0.45) display moderately diffuse attention, capturing scattered cellular structures, while smooth muscle (confidence 0.86, $\gamma = 0.15$) exhibits strong central focus reflecting organized tissue architecture. The background sample (confidence 0.90, $\gamma = 0.05$) generates distributed attention appropriate for uniform regions.

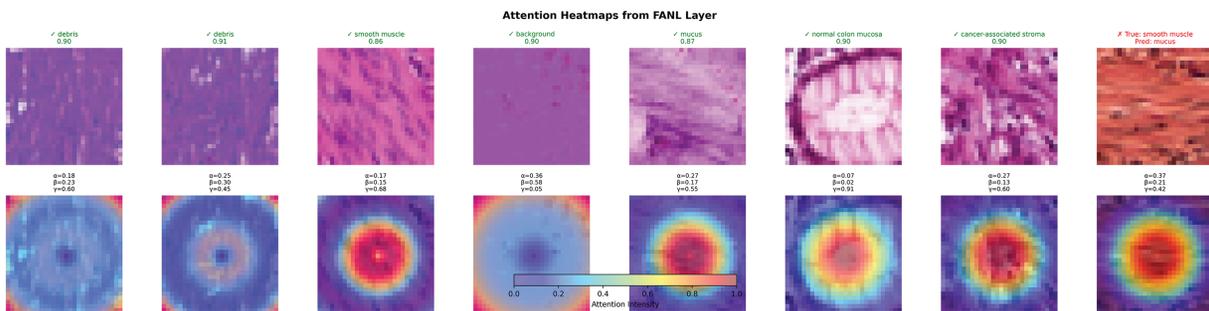


Figure 8: Adaptive patterns across different tissue types

Sample 8 shows a misclassification case: smooth muscle predicted as mucus with $\gamma = 0.21$ and moderately diffuse attention. Comparing this to correctly classified smooth muscle in sample 3 ($\gamma = 0.15$), the higher γ weight captures the morphological ambiguity between these tissue types. The elevated hesitation degree indicates the model recognizes the uncertainty despite incorrect prediction.

A clear pattern links fuzzy parameters to attention distribution: samples with high α weights (emphasizing membership μ) produce sharper attention peaks, while those with high γ weights generate broader coverage. This reflects the mechanism where the $\alpha \cdot \mu$ term concentrates attention on confident features, $\beta \cdot (1 - \nu)$ suppresses background, and $\gamma \cdot \pi$ distributes attention across uncertain regions.

4.4 Discussion

The proposed FANL model demonstrates superior performance across all comparative methods. On PathMNIST-30K, FANL achieves $84.41\% \pm 0.56\%$ test accuracy, substantially outperforming the baseline CNN ($82.72\% \pm 0.94\%$) by 1.69 percentage points. When scaled to the full PathMNIST dataset (107,180 samples), FANL maintains its advantage with $89.08\% \pm 0.23\%$ accuracy, improving upon CNN ($87.28\% \pm 0.47\%$) by 1.79 percentage points. The consistent performance gains across both dataset scales, coupled with reduced variance at larger sample sizes, demonstrate FANL’s robust generalization characteristics.

Interestingly, traditional attention mechanisms show limited or negative impact on performance. On PathMNIST-30K, SE-Net achieves $82.63\% \pm 0.65\%$ (0.09 points below baseline), CBAM reaches $82.89\% \pm 1.09\%$ (0.18 points above), CA attains $82.43\% \pm 1.14\%$ (0.29 points below), and ECA-Net drops to $81.86\% \pm 0.69\%$ (0.86 points below baseline). These patterns persist on the full dataset: SE-Net 87.16% (-0.13%), CBAM 87.64% ($+0.36\%$), ECA-Net 87.14% (-0.14%), CA 87.42% ($+0.14\%$). The consistent under-performance or marginal gains suggest that deterministic attention mechanisms may struggle to capture the nuanced uncertainty inherent in medical images, potentially over-emphasizing certain features while discarding subtle diagnostic cues.

FANL's improvements stem from three complementary design elements. The ternary fuzzy representation—comprising membership (μ), non-membership (ν), and hesitation (π) degrees—provides a richer vocabulary for modeling tissue boundary ambiguity and morphological uncertainty than binary attention weights can capture. The cross-layer attention guidance mechanism allows shallow-layer texture details to inform deeper semantic representations, creating more coherent feature hierarchies. Finally, the adaptive weight mechanism adjusts the influence of each fuzzy component based on image characteristics, eliminating the need for manual tuning across different pathological conditions.

Beyond accuracy improvements, FANL exhibits notably stable behavior. Its standard deviation ($\pm 0.56\%$ on 30 K samples, $\pm 0.23\%$ on the full dataset) is consistently lower than all baselines, suggesting reliable performance across different data samples and training runs. The computational cost is modest: FANL adds 10.97% more parameters (768.2 K total) and only 2.2% more FLOPs (6.00 M vs. CNN's 5.87 M). Other attention mechanisms introduce similar parameter overhead (SE: $+1.03\%$, CBAM: $+1.10\%$, CA: $+1.89\%$) but fail to deliver comparable gains, highlighting FANL's efficient use of additional capacity.

Evaluation of BloodMNIST provides insight into when fuzzy modeling is most beneficial. FANL achieves $96.21\% \pm 0.04\%$ accuracy, marginally below CBAM's $96.23\% \pm 0.26\%$ on this blood cell classification task. However, the baseline CNN already reaches 96.01% accuracy, suggesting the task is approaching saturation, where further improvements are difficult regardless of method. The key difference lies in task characteristics: blood cells have relatively consistent morphology within classes and clear boundaries between classes, whereas histopathological tissues exhibit significant ambiguity. FANL's larger gains on PathMNIST (1.79%) compared to BloodMNIST (0.20%) support the hypothesis that fuzzy uncertainty modeling is most valuable when dealing with genuinely ambiguous visual data.

4.5 Limitations and Future Directions

This study has several limitations that warrant discussion. First, the choice of the cluster number ($K = 3$) in FANL's fuzzy modeling module is manually predefined rather than adaptively determined. Although our ablation studies indicate that $K = 3$ offers an optimal balance between representational granularity and computational efficiency for the evaluated datasets, this fixed configuration may not generalize optimally across different imaging domains. Developing an adaptive mechanism to automatically determine K based on data characteristics would enhance FANL's flexibility, though this remains an open challenge in fuzzy clustering theory.

Second, while FANL's uncertainty quantification has been validated through selective prediction and calibration metrics, we have not conducted direct comparisons with other uncertainty-aware frameworks such as Bayesian CNNs or Monte Carlo Dropout. Incorporating such comparisons would provide stronger empirical evidence for the advantages of the proposed fuzzy-based approach.

Third, the current experiments are limited to microscopy images from histopathology and hematology. Whether FANL can generalize to radiological modalities (e.g., CT, MRI, X-ray)—which differ substantially in resolution, dimensionality, and noise distribution—remains an open question.

Fourth, the BloodMNIST results reveal that FANL's advantage becomes less pronounced on tasks where visual features are well-defined and the baseline model approaches performance saturation. This suggests that fuzzy uncertainty modeling provides the most benefit for inherently ambiguous or noisy visual data.

Finally, the present evaluation relies primarily on benchmark-level metrics rather than expert-based validation. For real-world deployment, a professional assessment would be necessary to determine whether FANL's improvements translate into meaningful diagnostic or interpretive utility. Addressing these limitations—through adaptive cluster selection, comprehensive uncertainty benchmarking, cross-modality generalization studies, and expert-in-the-loop validation—represents important directions for future research.

5 Conclusion

This paper presents a novel Fuzzy Attention Network Layer (FANL) that organically integrates intuitionistic fuzzy set theory with attention mechanisms in deep learning, providing a new solution pathway for addressing uncertainty in visual tasks. FANL employs a three-element fuzzy representation based on membership (μ), non-membership (ν), and hesitation (π) degrees to comprehensively characterize feature importance and uncertainty. Combined with adaptive fuzzy weight generation and cross-layer attention guidance mechanisms, FANL effectively extends and optimizes traditional attention methods.

Comprehensive experiments conducted on three benchmark datasets—PathMNIST-30000, full PathMNIST, and BloodMNIST—demonstrate the effectiveness and generalizability of the proposed Fuzzy Attention Network Layer (FANL). Compared with the standard CNN baseline, FANL achieves consistent performance gains across all datasets, including a 1.07% accuracy improvement on full PathMNIST ($84.97\% \pm 0.32\%$) and stable macro-average F1-scores of 0.86 and 0.96 on PathMNIST-30000 and BloodMNIST, respectively. Notably, FANL is the only attention mechanism that consistently outperforms the baseline CNN, whereas all five conventional attention modules (SE-Net, CBAM, ECA-Net, CA, and Coordinate Attention) exhibit accuracy degradation or instability across datasets.

Finally, the adaptive weight fusion strategy automatically learns optimal contributions of the three uncertainty measures (μ, ν, π) without manual tuning. This design maintains lightweight computation while providing interpretable and uncertainty-aware attention modeling across diverse feature distributions, making FANL suitable for practical medical imaging applications with varying data characteristics.

The successful application of FANL demonstrates the broad prospects of combining fuzzy logic theory with modern deep learning techniques. This approach not only theoretically extends the expressive capability of attention mechanisms but also practically exhibits unique advantages in handling ambiguous boundaries and representational uncertainty in medical imaging. The integration of intuitionistic fuzzy set theory provides a principled framework for modeling the inherent uncertainty in pathological image analysis, offering valuable insights for the broader medical AI community. Recent advances in fuzzy-attention integration further support this promising research direction. Wang et al. [50] demonstrated fuzzy hierarchical fusion attention for medical image super-resolution, while Nan et al. [13] developed fuzzy attention networks for airway segmentation discontinuity challenges. These developments highlight the growing trend of combining fuzzy logic with attention mechanisms across diverse medical imaging applications. Future work will explore FANL's application potential in more complex visual tasks, investigate integration with other

fuzzy-attention approaches, and focus on developing more efficient fuzzy modeling strategies to advance the continuous development of uncertainty-aware deep learning methods.

Acknowledgement: Not applicable.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: conceptualization, Zheng Zhao and Kwang Baek Kim; methodology, Zheng Zhao and Doo Heon Song; investigation, Zheng Zhao and Kwang Baek Kim; writing—original draft preparation, Zheng Zhao, Doo Heon Song and Kwang Baek Kim; writing—review and editing, Zheng Zhao, Kwang Baek Kim and Doo Heon Song; visualization, Zheng Zhao; project administration, Kwang Baek Kim; funding acquisition, Kwang Baek Kim. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: All datasets used in this study are publicly available. PathMNIST (89,996 samples, 9 classes) and BloodMNIST (17,092 samples, 8 classes) can be accessed through the MedMNIST repository (<https://medmnist.com/>). To facilitate full reproducibility, we release the exact training/validation/testing indices used in all experiments. A compiled reproducibility package is provided, containing: (1) the pre-trained FANL model in TorchScript format; (2) minimal configuration files specifying data preprocessing, training settings, and model-export parameters; and (3) minimal runnable scripts for model loading and GPU-accelerated inference. Due to ongoing intellectual property and commercialization considerations, the complete training code will be released after the current project cycle. All reproducibility materials are publicly available at: <https://github.com/ZhaoDreamGo/FANL-Reproducibility-Package>. All materials will remain accessible upon publication.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Alzubaidi L, Bai J, Al-Sabaawi A, Santamaría J, Albahri AS, Al-dabbagh BSN, et al. A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. *J Big Data*. 2023;10(1):46. doi:10.1186/s40537-023-00727-2.
2. Brauwiers G, Frasincar F. A general survey on attention mechanisms in deep learning. *IEEE Trans Knowl Data Eng*. 2023;35(4):3279–98. doi:10.1109/tkde.2021.3126456.
3. Guo MH, Xu TX, Liu JJ, Liu ZN, Jiang PT, Mu TJ, et al. Attention mechanisms in computer vision: a survey. *Comp Visual Med*. 2022;8(3):331–68. doi:10.1007/s41095-022-0271-y.
4. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2018 Jun 18–23; Salt Lake City, UT, USA. p. 7132–41.
5. Wang X, Girshick R, Gupta A, He K. Non-local neural networks. In: *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2018 Jun 18–23; Salt Lake City, UT, USA. p. 7794–803. doi:10.1109/cvpr.2018.00813.
6. Woo S, Park J, Lee JY, Kweon IS. CBAM: convolutional block attention module. In: *Proceedings of the Computer Vision—ECCV 2018*; 2018 Sep 8–14; Munich, Germany. Cham, Switzerland: Springer International Publishing; 2018. p. 3–19. doi:10.1007/978-3-030-01234-2_1.
7. Park J, Woo S, Lee JY, Kweon IS. BAM: bottleneck attention module. *arXiv:1807.06514*. 2018.
8. Wang F, Jiang M, Qian C, Yang S, Li C, Zhang H, et al. Residual attention network for image classification. In: *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2017 Jul 21–26; Honolulu, HI, USA. p. 6450–8. doi:10.1109/cvpr.2017.683.

9. Abdar M, Pourpanah F, Hussain S, Rezazadegan D, Liu L, Ghavamzadeh M, et al. A review of uncertainty quantification in deep learning: techniques, applications and challenges. *Inf Fusion*. 2021;76(1):243–97. doi:10.1016/j.inffus.2021.05.008.
10. Zadeh LA, Klir GJ, Yuan B. *Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers*. Singapore: World Scientific; 1996.
11. Kar S, Das S, Ghosh PK. Applications of neuro fuzzy systems: a brief review and future outline. *Appl Soft Comput*. 2014;15:243–59. doi:10.1016/j.asoc.2013.10.014.
12. Atanassov KT. Intuitionistic fuzzy sets. *Fuzzy Sets Syst*. 1986;20(1):87–96. doi:10.1016/S0165-0114(86)80034-3.
13. Nan Y, Del Ser J, Tang Z, Tang P, Xing X, Fang Y, et al. Fuzzy attention neural network to tackle discontinuity in airway segmentation. *IEEE Trans Neural Netw Learning Syst*. 2024;35(6):7391–404. doi:10.1109/tnnls.2023.3269223.
14. Yang J, Shi R, Wei D, Liu Z, Zhao L, Ke B, et al. MedMNIST v2—a large-scale lightweight benchmark for 2D and 3D biomedical image classification. *Sci Data*. 2023;10(1):41. doi:10.1038/s41597-022-01721-8.
15. Wang Q, Wu B, Zhu P, Li P, Zuo W, Hu Q. ECA-net: efficient channel attention for deep convolutional neural networks. In: *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun 13–19; Seattle, WA, USA*. p. 11531–9. doi:10.1109/cvpr42600.2020.01155.
16. Wang Y, Wang W, Li Y, Jia Y, Xu Y, Ling Y, et al. An attention mechanism module with spatial perception and channel information interaction. *Complex Intell Syst*. 2024;10(4):5427–44. doi:10.1007/s40747-024-01445-9.
17. Xu W, Wan Y, Zhao D. SFA: efficient attention mechanism for superior CNN performance. *Neural Process Lett*. 2025;57(2):38. doi:10.1007/s11063-025-11748-8.
18. Jaderberg M, Simonyan K, Zisserman A, Kavukcuoglu K. Spatial transformer networks. *Adv Neural Inf Process Syst*. 2015;28:1–9. doi:10.5555/2969442.2969465.
19. Hong J, Lee B, Ko K, Ko H. Fast non-local attention network for light super-resolution. *J Vis Commun Image Represent*. 2023;95:103861. doi:10.1016/j.jvcir.2023.103861.
20. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. UNet++: redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Trans Med Imaging*. 2020;39(6):1856–67. doi:10.1109/TMI.2019.2959609.
21. Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design. In: *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2021 Jun 20–25; Nashville, TN, USA*. p. 13708–17. doi:10.1109/cvpr46437.2021.01350.
22. Rajafillah C, El Moutaouakil K, Patriciu AM, Yahyaouy A, Riffi J. INT-FUP: intuitionistic fuzzy pooling. *Mathematics*. 2024;12(11):1740. doi:10.3390/math12111740.
23. Atanassov K, Sotirov S, Pencheva T. Intuitionistic fuzzy deep neural network. *Mathematics*. 2023;11(3):716. doi:10.3390/math11030716.
24. Zhang S, Fang Y, Nan Y, Wang S, Ding W, Ong YS, et al. Fuzzy attention-based border rendering network for lung organ segmentation. In: *Proceedings of the 27th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI); 2024 Oct 6–10; Marrakesh, Morocco*.
25. Lin CT, Lee CSG. Neural-network-based fuzzy logic control and decision system. *IEEE Trans Comput*. 1991;40(12):1320–36. doi:10.1109/12.106218.
26. Nauck D, Klawonn F, Kruse R. *Foundations of neuro-fuzzy systems*. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 1997.
27. Havens TC, Bezdek JC, Leckie C, Hall LO, Palaniswami M. Fuzzy c-means algorithms for very large data. *IEEE Trans Fuzzy Syst*. 2012;20(6):1130–46. doi:10.1109/tfuzz.2012.2201485.
28. Despotović I, Goossens B, Philips W. MRI segmentation of the human brain: challenges, methods, and applications. *Comput Math Methods Med*. 2015;2015:450341. doi:10.1155/2015/450341.
29. Siddique N, Paheding S, Elkin CP, Devabhaktuni V. U-Net and its variants for medical image segmentation: a review of theory and applications. *IEEE Access*. 2021;9:82031–57. doi:10.1109/ACCESS.2021.3086020.
30. Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. *Proc Mach Learn Res*. 2016;48:1050–9.

31. Kendall A, Gal Y. What uncertainties do we need in Bayesian deep learning for computer vision? *Adv Neural Inf Process Syst.* 2017;30:5574–84.
32. Sensoy M, Kaplan L, Kandemir M. Evidential deep learning to quantify classification uncertainty. *Adv Neural Inf Process Syst.* 2018;31:3179–89.
33. Amini A, Schwarting W, Soleimany A, Rus D. Deep evidential regression. *Adv Neural Inf Process Syst.* 2020;33:14927–37.
34. Xu Z. Intuitionistic fuzzy aggregation operators. *IEEE Trans Fuzzy Syst.* 2007;15(6):1179–87. doi:10.1109/tfuzz.2006.890678.
35. Zheng Y, Xu Z, Wu T, Yi Z. A systematic survey of fuzzy deep learning for uncertain medical data. *Artif Intell Rev.* 2024;57(9):230. doi:10.1007/s10462-024-10871-7.
36. Bezdek JC. *Pattern recognition with fuzzy objective function algorithms.* New York, NY, USA: Plenum Press; 1981.
37. Chaira T. A novel intuitionistic fuzzy C means clustering algorithm and its application to medical images. *Appl Soft Comput.* 2011;11(2):1711–7. doi:10.1016/j.asoc.2010.05.005.
38. Memiş S, Şola Erduran F, Aydoğan H. Adaptive machine learning approaches utilizing soft decision-making via intuitionistic fuzzy parameterized intuitionistic fuzzy soft matrices. *PeerJ Comput Sci.* 2025;11:e2703. doi:10.7717/peerj-cs.2703.
39. Marsala C. Building intuitionistic fuzzy sets in machine learning. In: *Proceedings of the 13th International Workshop on Fuzzy Logic and Applications (WILF 2021); 2021 Dec 20–22; Vietri sul Mare, Italy.*
40. Tirupal T, Chandra Mohan B, Srinivas Kumar S. Multimodal medical image fusion based on interval-valued intuitionistic fuzzy sets. In: *Machines, mechanism and robotics.* Singapore: Springer Singapore; 2021. p. 965–71 doi: 10.1007/978-981-16-0550-5_91.
41. Begoli E, Bhattacharya T, Kusnezov D. The need for uncertainty quantification in machine-assisted medical decision making. *Nat Mach Intell.* 2019;1(1):20–3. doi:10.1038/s42256-018-0004-1.
42. Lambert B, Forbes F, Doyle S, Dehaene H, Dojat M. Trustworthy clinical AI solutions: a unified review of uncertainty quantification in deep learning models for medical image analysis. *Artif Intell Med.* 2024;150:102830. doi:10.1016/j.artmed.2024.102830.
43. Lakshminarayanan B, Pritzel A, Blundell C. Simple and scalable predictive uncertainty estimation using deep ensembles. *Adv Neural Inf Process Syst.* 2017;30:6402–13. doi:10.5555/3295222.3295387.
44. Guo C, Pleiss G, Sun Y, Weinberger KQ. On calibration of modern neural networks. *Proc Mach Learn Res.* 2017;70:1321–30. doi:10.5555/3305381.3305518.
45. Lin TY, Dollar P, Girshick R, He K, Hariharan B, Szeliski S. Feature pyramid networks for object detection. In: *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017 Jul 21–26; Honolulu, HI, USA.* p. 936–44.
46. Liu S, Qi L, Qin H, Shi J, Jia J. Path aggregation network for instance segmentation. In: *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2018 Jun 18–23; Salt Lake City, UT, USA.* p. 8759–68. doi:10.1109/cvpr.2018.00913.
47. Li H, Xiong P, An J, Wang L. Pyramid attention network for semantic segmentation. arXiv:1805.10180. 2018.
48. Zagoruyko S, Komodakis N. Paying more attention to attention: improving the performance of convolutional neural networks via attention transfer. arXiv:1612.03928. 2016.
49. Loshchilov I, Hutter F. Decoupled weight decay regularization 2019. [cited 2025 Jan 1]. Available from: <https://openreview.net/forum?id=Bkg6RiCqY7>.
50. Wang C, Lv X, Shao M, Qian Y, Zhang Y. A novel fuzzy hierarchical fusion attention convolution neural network for medical image super-resolution reconstruction. *Inf Sci.* 2023;622:424–36. doi:10.1016/j.ins.2022.11.140.