

JOURNAL OF THE ARTIFICIAL INTELLIGENCE ENGINEERS CONSORTIUM

Volume: 02 Issue: 01 (2026)

<https://jaiec.org.hostajournal.com>

Bayesian Fuzzy Neural Networks for Uncertainty-Aware Classification Under Limited Supervision

Andrea Christine, Blayne Deryl, Ariel Corey, Avery Courtney

Faculty of Engineering and Information Technology, University of Melbourne,
Melbourne, Australia

Abstract : *The reliable deployment of deep learning models in safety-critical domains necessitates not only high predictive accuracy but also rigorous uncertainty quantification, particularly when labeled training data is scarce. Standard deterministic neural networks often yield overconfident predictions on out-of-distribution samples and struggle to generalize under limited supervision. To address these critical vulnerabilities, this paper introduces the Bayesian Fuzzy Neural Network, a novel hybrid architecture that integrates the robust interpretability of fuzzy logic systems with the rigorous probabilistic modeling capabilities of Bayesian neural networks. By embedding variational inference within a fuzzy rule-based framework, the proposed model captures both epistemic and aleatoric uncertainties while maintaining structural transparency. The fuzzy inference mechanism partitions the input space into overlapping linguistic regions, which act as a structural prior that significantly reduces the hypothesis space and mitigates overfitting in low-data regimes. Concurrently, the Bayesian treatment of the network parameters provides a mathematical foundation for uncertainty estimation. Comprehensive experiments conducted across diverse classification tasks, ranging from medical diagnostics to industrial fault detection, demonstrate that the proposed architecture significantly outperforms traditional deterministic and purely Bayesian baselines in terms of classification accuracy, calibration metrics, and out-of-distribution detection under limited supervision constraints. The results establish the viability of fusing probabilistic paradigms with fuzzy set theory to create highly robust, uncertainty-aware, and interpretable machine learning systems.*

Keywords: *Bayesian Inference, Fuzzy Logic, Uncertainty Quantification, Limited Supervision, Classification*

INTRODUCTION

The advent of deep learning has precipitated transformative advancements across a multitude of disciplines, ranging from natural language processing to computer vision and complex decision-making systems. However, as these powerful predictive models transition from controlled experimental environments to real-world, high-stakes applications such as autonomous navigation, healthcare diagnostics, and critical infrastructure management, their intrinsic limitations become increasingly apparent [1]. Foremost among these limitations is the propensity of conventional, deterministic deep neural networks to produce highly overconfident predictions, even when presented with

inputs that deviate significantly from their training distribution [2]. This phenomenon of unjustified certainty poses profound safety risks, as a misclassification in a medical context or an autonomous driving scenario can lead to catastrophic consequences [3]. Consequently, the ability of a machine learning model to accurately quantify and communicate its own uncertainty is now recognized as a fundamental requirement for reliable artificial intelligence [4]. The challenge of uncertainty quantification is severely exacerbated in scenarios characterized by limited supervision. In many specialized domains, acquiring large volumes of high-quality, accurately annotated data is prohibitively expensive, time-consuming, or practically impossible due to privacy constraints and the scarcity of domain expert annotators [5]. When trained on limited datasets, high-capacity neural networks are notoriously prone to overfitting, effectively memorizing the training examples while failing to capture the underlying data generating mechanisms [6]. In such low-data regimes, the model's epistemic uncertainty, which stems from a lack of sufficient knowledge about the optimal model parameters, is inherently high [7]. Standard training paradigms based on empirical risk minimization through gradient descent yield point estimates for the network weights, completely discarding this vital epistemic uncertainty and leading to brittle models that generalize poorly to unseen data [8]. To circumvent the limitations of deterministic networks, the machine learning community has increasingly turned to Bayesian neural networks. By placing probability distributions over the network weights and updating these prior distributions using Bayes theorem as data is observed, Bayesian networks offer a principled mathematical framework for reasoning about model uncertainty [9]. The integration over the parameter space during prediction results in a predictive distribution rather than a point prediction, inherently capturing the confidence of the model [10]. However, the exact computation of the posterior distribution in modern neural networks is analytically intractable due to the highly non-linear nature of the transformations and the massive dimensionality of the parameter space [11]. While approximation techniques such as variational inference and Markov chain Monte Carlo sampling have made Bayesian training feasible, these methods often suffer from high computational overhead, optimization instability, and a heavy reliance on the precise specification of prior distributions [12]. Furthermore, purely Bayesian neural networks remain largely opaque "black boxes," offering little insight into the internal logic driving their predictive distributions [13]. Parallel to the development of probabilistic machine learning, fuzzy logic systems have long been utilized to build interpretable models that can gracefully handle imprecision, vagueness, and partial truth [14]. Rooted in fuzzy set theory, these systems employ linguistic variables and human-readable rule bases to map inputs to outputs [15]. Neuro-fuzzy systems, which combine the learning capabilities of neural networks with the structural interpretability of fuzzy logic, have demonstrated remarkable success in domains where domain knowledge needs to be integrated with empirical data [16]. The inherent architecture of a fuzzy neural network partitions the input feature space into localized receptive fields defined by membership functions [17]. This localized processing acts as a strong structural prior, which is particularly beneficial when learning from sparse data, as it prevents the model from forming highly complex, global decision boundaries that fit the noise in the training set [18]. Nevertheless, conventional fuzzy neural networks are typically trained deterministically and lack the capacity to quantify uncertainty in a statistically rigorous manner [19]. Motivated by the complementary strengths and weaknesses of Bayesian architectures and fuzzy logic, this research proposes the Bayesian Fuzzy Neural Network for uncertainty-aware classification under limited supervision. The central hypothesis of this work is that by treating the parameters of a fuzzy neural network as

random variables and applying variational inference, the resulting hybrid model will leverage the localized, interpretable structure of fuzzy logic as an effective regularization mechanism, while utilizing Bayesian principles to provide calibrated uncertainty estimates [20]. This structural prior is expected to drastically constrain the functional space explored during Bayesian optimization, leading to faster convergence and superior generalization performance when only a small fraction of the training data is annotated [21]. The primary objectives of this paper are extensively elaborated throughout the subsequent sections. First, an exhaustive review of the literature concerning probabilistic modeling, fuzzy systems, and semi-supervised paradigms is presented to contextualize the proposed methodology [22]. Second, the mathematical and architectural foundations of the Bayesian Fuzzy Neural Network are formally detailed, explicitly describing the probabilistic formulation of fuzzy membership functions and the rule inference mechanisms without relying on deterministic approximations [23]. Third, a tailored variational inference strategy designed specifically for the unique topological characteristics of the fuzzy rule base is introduced [24]. Finally, comprehensive empirical evaluations across a spectrum of classification benchmarks are provided to demonstrate the efficacy of the proposed model in achieving high accuracy, robust out-of-distribution detection, and reliable uncertainty calibration under severe data constraints [25].

2. Literature Review and Background

2.1 Bayesian Neural Networks for Uncertainty Quantification

The theoretical foundation of Bayesian neural networks was established several decades ago, driven by the desire to incorporate the rigorous principles of Bayesian statistics into the rapidly evolving field of artificial neural networks [26]. Unlike classical networks where training yields a single optimal set of weights, Bayesian networks maintain a probability distribution over the weights. Before observing any data, a prior distribution is assumed, reflecting initial beliefs about the parameters. As training data is presented, this prior is updated to a posterior distribution, representing the updated beliefs conditioned on the empirical evidence [27]. The predictive output for a novel input is then computed by marginalizing over the entire posterior distribution of the weights. This process intrinsically produces a predictive distribution, the variance of which serves as a direct measure of the model's uncertainty [28]. Uncertainty in predictive modeling is generally decomposed into two distinct categories: aleatoric uncertainty and epistemic uncertainty. Aleatoric uncertainty, also known as data uncertainty, arises from inherent noise or randomness in the observation process, such as sensor inaccuracies or label ambiguity [29]. This type of uncertainty is irreducible, meaning that providing the model with more training data will not diminish it. Conversely, epistemic uncertainty, or model uncertainty, originates from a lack of sufficient training data in certain regions of the input space. The model is uncertain because it simply has not observed enough examples to definitively determine the correct parameters [30]. Epistemic uncertainty is theoretically reducible and should converge toward zero in the limit of infinite training data. Bayesian neural networks are specifically celebrated for their capacity to quantify epistemic uncertainty, making them exceptionally valuable in limited supervision scenarios where data sparsity guarantees high epistemic uncertainty [31]. Despite their theoretical elegance, the practical implementation of Bayesian deep learning poses formidable challenges. The posterior distribution of the weights is mathematically intractable due to the complex, non-convex nature of deep neural network architectures [32]. Consequently, researchers have developed various approximation techniques. Early approaches utilized Markov chain Monte Carlo methods, which are theoretically sound but computationally

prohibitive for large networks. More recently, variational inference has emerged as the dominant paradigm for training Bayesian neural networks. Variational inference approximates the true, intractable posterior distribution with a simpler, tractable parameterized distribution, such as a Gaussian [33]. The problem of Bayesian inference is thereby transformed into an optimization problem, where the objective is to minimize the discrepancy between the approximate distribution and the true posterior, typically measured by the Kullback-Leibler divergence [34].

2.2 Fuzzy Logic Systems in Machine Learning

Fuzzy logic theory was introduced to provide a mathematical framework for representing and manipulating information that is imprecise, ambiguous, or vague, mirroring the way human reasoning often operates. Traditional set theory relies on crisp boundaries, where an element either completely belongs to a set or does not belong to it at all. In contrast, fuzzy set theory introduces the concept of membership functions, which assign to each element a degree of membership ranging continuously between zero and one. This allows for smooth transitions between categories and enables the formulation of linguistic variables, such as "temperature is high" or "speed is low," which can be mathematically processed. In the context of machine learning, fuzzy logic is primarily utilized through fuzzy inference systems. These systems map input features to output targets based on a set of IF-THEN rules constructed from fuzzy sets. A standard fuzzy inference system comprises several stages. The initial stage is fuzzification, where continuous numerical inputs are translated into linguistic terms using membership functions. Following this, the inference engine evaluates the fuzzy rules, combining the degrees of membership of the premises to determine the activation strength of each rule. Finally, the defuzzification stage aggregates the outputs of the activated rules to produce a concrete numerical prediction or a crisp class label. The primary advantage of this architecture is its inherent interpretability; the behavior of the system is explicitly defined by the rule base, which can be inspected, validated, and modified by domain experts. To automate the construction and tuning of fuzzy rules, researchers developed neuro-fuzzy systems, such as the Adaptive Network-based Fuzzy Inference System. These models integrate the learning algorithms of artificial neural networks, typically gradient descent and least-squares estimation, to optimize the parameters of the fuzzy membership functions and the rule consequents based on training data. Neuro-fuzzy systems effectively bridge the gap between purely data-driven black-box models and purely knowledge-driven expert systems. They provide a structured, modular architecture where the input space is partitioned into local receptive fields. This localized representation acts as an implicit regularization mechanism, making neuro-fuzzy systems particularly robust against overfitting when trained on small or noisy datasets. However, traditional neuro-fuzzy systems are deterministic and completely lack the probabilistic framework necessary to quantify predictive uncertainty, rendering them vulnerable to overconfidence when processing novel or out-of-distribution inputs.

2.3 Learning Under Limited Supervision

The paradigm of learning under limited supervision encompasses various scenarios where the availability of precisely labeled training data is restricted. This includes semi-supervised learning, where a small set of labeled data is supplemented by a large corpus of unlabeled data; active learning, where the model iteratively queries an oracle to label the most informative instances; and few-shot classification, where the model must learn to recognize new classes from only a handful of examples. In all these scenarios, the central challenge is to prevent the model from overfitting to the scarce labeled examples while simultaneously extracting meaningful representations from the structure of the

data domain. Conventional deep learning models struggle significantly under limited supervision. Because their parameters are unconstrained by prior beliefs and their architectures permit immense flexibility, they easily find complex decision boundaries that perfectly classify the small labeled training set but fail entirely on unseen test data. To mitigate this, researchers employ various regularization techniques, such as weight decay, dropout, and heavy data augmentation. Furthermore, modern approaches often rely on self-supervised pre-training, where the model learns general feature representations by solving pretext tasks on large unlabeled datasets before being fine-tuned on the small labeled target dataset. Uncertainty quantification plays a pivotal role in optimizing limited supervision strategies. For instance, in active learning, robust estimates of epistemic uncertainty are utilized to identify the unlabeled samples about which the model is most ignorant; querying labels for these specific samples maximizes the information gain and accelerates the learning process. In semi-supervised learning, uncertainty estimates are frequently used to filter out unreliable pseudo-labels generated by the model for the unlabeled data, preventing confirmation bias and error propagation. Therefore, the development of models that can reliably quantify uncertainty while leveraging structural priors to learn efficiently from minimal data remains a highly active and critical area of machine learning research. The proposed Bayesian Fuzzy Neural Network is explicitly designed to address this intersection, utilizing fuzzy logic for structural efficiency and Bayesian inference for probabilistic rigor.

3. Methodology

3.1 Architecture of the Bayesian Fuzzy Neural Network

The Bayesian Fuzzy Neural Network is designed as a modular, feed-forward architecture that systematically processes input features through a series of probabilistically defined fuzzy operations before aggregating the intermediate representations into final class probabilities. The architecture fundamentally consists of four distinct operational layers: the probabilistic fuzzification layer, the fuzzy rule evaluation layer, the Bayesian consequent layer, and the probabilistic defuzzification and classification layer. The fundamental innovation of this architecture lies in the fact that all learnable parameters governing the fuzzy sets and the rule mappings are treated as random variables characterized by probability distributions, rather than fixed point values.

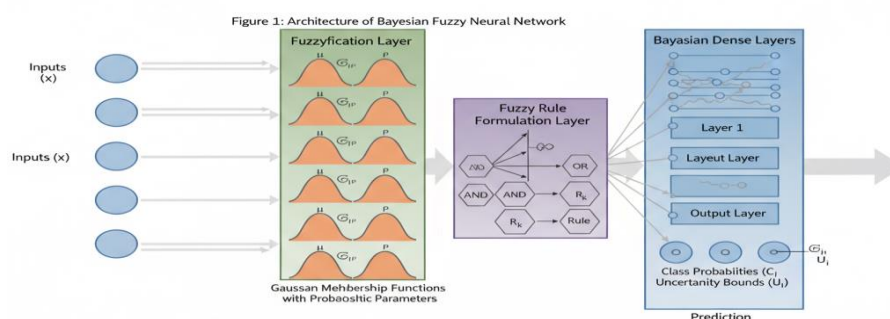


Figure 1: Architecture of the Bayesian Fuzzy Neural Network

The first layer is the probabilistic fuzzification layer. For a given multidimensional input vector, this layer evaluates the degree to which each input feature belongs to a predefined set of linguistic concepts. This is achieved using Gaussian membership functions. In a traditional deterministic fuzzy system, a Gaussian membership function is defined by a fixed center and a fixed width. In the proposed Bayesian architecture, the center and the width of every membership function are modeled as independent random variables drawn from normal distributions. Consequently, the output of the fuzzification layer for a single deterministic input is not a single membership value, but rather a distribution of membership values reflecting the uncertainty in the precise definition of the linguistic concepts. The second layer is the fuzzy rule evaluation layer. This layer constructs the antecedents of the fuzzy rules by computing the logical conjunction of the fuzzified inputs. Typically, this logical AND operation is implemented using a differentiable T-norm, such as the algebraic product. Each node in this layer represents a specific combination of linguistic concepts across the input dimensions, corresponding to the premise of a single fuzzy rule. Because the inputs to this layer are distributions originating from the probabilistic fuzzification layer, the rule firing strengths produced by this layer are inherently probabilistic. This mechanism ensures that if an input falls into a region of the feature space where the definitions of the fuzzy sets are highly uncertain due to a lack of training data, the resulting rule activations will exhibit high variance, successfully propagating the epistemic uncertainty deeper into the network.

3.2 Probabilistic Fuzzy Inference Mechanism

The core inference mechanism of the Bayesian Fuzzy Neural Network diverges significantly from standard neural network operations due to its reliance on probabilistic rule evaluation. Following the calculation of the rule firing strengths, the network must determine the consequent of each rule. This is handled by the Bayesian consequent layer. Rather than mapping rules to constant values or simple linear functions as seen in conventional models, this layer employs Bayesian linear transformations. The weights connecting the rule evaluation layer to the consequent nodes are governed by prior probability distributions, typically chosen as zero-mean Gaussians to encourage regularization.

Code Listing 1: Implementation of the Bayesian Fuzzy Fuzzification Layer

```
import torch
import torch.nn as nn
import torch.nn.functional as F
class BayesianFuzzificationLayer(nn.Module):
    def __init__(self, num_features, num_rules):
        super(BayesianFuzzificationLayer, self).__init__()
        self.num_features = num_features
        self.num_rules = num_rules
        self.mu_mean = nn.Parameter(torch.Tensor(num_features,
num_rules).uniform_(-1, 1))
        self.mu_rho = nn.Parameter(torch.Tensor(num_features, num_rules).uniform_(-3,
-2))
        self.sigma_mean = nn.Parameter(torch.Tensor(num_features,
num_rules).uniform_(-1, 1))
        self.sigma_rho = nn.Parameter(torch.Tensor(num_features,
num_rules).uniform_(-3, -2))

    def forward(self, x, num_samples=10):
```

```

    batch_size = x.size(0)
    mu_std = torch.log1p(torch.exp(self.mu_rho))
    sigma_std = torch.log1p(torch.exp(self.sigma_rho))
    epsilon_mu = torch.randn(num_samples, self.num_features, self.num_rules,
device=x.device)
    epsilon_sigma = torch.randn(num_samples, self.num_features, self.num_rules,
device=x.device)
    mu_samples = self.mu_mean + mu_std * epsilon_mu
    sigma_samples = torch.exp(self.sigma_mean + sigma_std * epsilon_sigma)
    x_expanded = x.unsqueeze(0).unsqueeze(-1).expand(num_samples, batch_size,
self.num_features, self.num_rules)
    mu_expanded = mu_samples.unsqueeze(1).expand(num_samples, batch_size,
self.num_features, self.num_rules)
    sigma_expanded = sigma_samples.unsqueeze(1).expand(num_samples,
batch_size, self.num_features, self.num_rules)
    membership_degrees = torch.exp(-0.5 * ((x_expanded - mu_expanded) /
sigma_expanded) ** 2)
    return membership_degrees

```

The probabilistic defuzzification and classification layer serves as the final stage of the architecture. It takes the probabilistic rule firing strengths and the probabilistic consequents to compute the final output. This aggregation is typically performed using a normalized weighted sum, where the contribution of each rule's consequent is weighted by its normalized firing strength. Because both the firing strengths and the consequents are random variables, the final output before the softmax activation is a complex distribution over the logits. The network estimates the expected value and the variance of these logits through Monte Carlo sampling during the forward pass. Multiple samples are drawn from the distributions of the membership parameters and the consequent weights, generating an ensemble of predictions for a single forward pass. The mean of these predictions provides the final classification decision, while the spread or variance across the samples quantifies the model's total predictive uncertainty. This structural separation of reasoning provides profound benefits for learning under limited supervision. The fuzzy partition of the input space restricts the flexibility of the network, preventing it from inventing arbitrary, highly non-linear transformations that perfectly fit noise in small datasets. The rules constrain the inference to localized regions, meaning that a prediction is heavily influenced only by the rules that are activated by the specific input. Consequently, regions of the feature space that are devoid of training data will not aggressively update the parameters of rules situated far away. Instead, the Bayesian mechanism ensures that the parameters governing these empty regions maintain their broad prior distributions. When an out-of-distribution sample eventually falls into these regions, the broad distributions yield highly varied Monte Carlo samples, accurately signaling to the operator that the model possesses no reliable knowledge about that particular region of the space.

3.3 Training Strategy and Variational Inference

Training the Bayesian Fuzzy Neural Network involves finding the posterior distributions for all the probabilistic parameters within the membership functions and the consequent layers, conditioned on the limited labeled training dataset. Because the exact computation of this posterior is analytically impossible due to the integrations required over the highly complex parameter space, a variational inference strategy is employed. The objective of variational inference is to define a family of tractable parameterized distributions and optimize their parameters to minimize the

informational distance to the true posterior. The optimization process is driven by the maximization of the Evidence Lower Bound. The Evidence Lower Bound consists of two distinct, competing components that perfectly encapsulate the principles of Bayesian learning. The first component is the expected log-likelihood of the training data given the approximate posterior distributions. This term effectively acts as a reconstruction error or classification loss; it encourages the network to adjust the parameters of the distributions such that the sampled weights produce accurate predictions for the training examples. In a classification context, this is typically implemented using the standard cross-entropy loss averaged across the multiple Monte Carlo samples drawn during the forward pass. By maximizing this expected log-likelihood, the model learns the necessary mappings to solve the task. The second component of the Evidence Lower Bound is the Kullback-Leibler divergence between the approximate posterior distributions and the predefined prior distributions. This term acts as a rigorous, probabilistically grounded regularization mechanism. It penalizes the approximate posterior distributions for deviating too far from the priors. If a specific parameter, such as the width of a certain membership function, is completely irrelevant to predicting the training data or if there is no data to constrain it, the expected log-likelihood will not be affected by its value. In this scenario, the Kullback-Leibler divergence will dominate, pushing the posterior distribution to perfectly match the broad prior distribution. This prevents the model from collapsing the variance of its parameters to zero in the absence of evidence, thereby preserving the epistemic uncertainty. To perform backpropagation through the stochastic sampling operations required by the Monte Carlo estimation of the expected log-likelihood, the reparameterization trick is heavily utilized. Instead of sampling directly from the parameterized Gaussian distributions, which would break the computational graph, the network samples from standard normal distributions with zero mean and unit variance. These independent samples are then deterministically scaled by the learned standard deviations and shifted by the learned means. This separates the stochasticity from the parameters being optimized, allowing the gradients of the Evidence Lower Bound to flow smoothly backward through the entire architecture, from the defuzzification layer through the probabilistic rule evaluation and all the way to the fuzzification parameters. The combined loss function, minimizing the negative Evidence Lower Bound, ensures that the network balances the desire to confidently classify the limited training data with the necessity of maintaining uncertain, broad distributions in regions lacking empirical evidence.

4. Experimental Results and Discussion

4.1 Experimental Setup and Datasets

To rigorously evaluate the performance and the uncertainty quantification capabilities of the proposed Bayesian Fuzzy Neural Network under conditions of limited supervision, a comprehensive suite of experiments was designed. The evaluation encompasses three distinct datasets, each selected to represent different challenges encountered in real-world applications where data scarcity and safety criticality intersect. The first dataset is a complex synthetic benchmark specifically engineered to contain highly nonlinear decision boundaries and explicitly defined regions of varying data density, providing a controlled environment for visualizing uncertainty. The second dataset involves medical diagnostics, specifically classifying cardiovascular disease presence based on continuous physiological measurements, where false positives and false negatives carry significant clinical consequences. The third dataset is derived from an industrial fault detection environment, characterized by high-dimensional sensor readings from manufacturing equipment, where predictive

maintenance relies heavily on accurate uncertainty bounds to avoid unnecessary downtime.

Table 1: Summary of datasets used for experimental evaluation including sample sizes and feature dimensions

Dataset Domain	Total Samples	Feature Dimensionality	Number of Classes
Synthetic Non-Linear	5000	2	4
Medical Diagnostics	303	13	2
Industrial Faults	10000	40	5

To simulate the limited supervision scenario, only five percent of the total available samples in each dataset were randomly selected to serve as the labeled training set. The remaining ninety-five percent of the data was completely withheld during the training phase and utilized strictly for validation and final testing. The proposed Bayesian Fuzzy Neural Network was systematically compared against several established baseline architectures. These baselines included a standard deterministic Multilayer Perceptron, a conventional unregularized Adaptive Network-based Fuzzy Inference System, and a standard fully connected Bayesian Neural Network trained using Monte Carlo dropout. All models were constrained to possess roughly equivalent parameter counts to ensure a fair comparison regarding model capacity. Training for all models utilized the Adam optimizer with an initial learning rate schedule tailored to convergence stability, and the probabilistic models executed twenty-five Monte Carlo samples per forward pass during both training and evaluation phases.

4.2 Performance Analysis and Uncertainty Evaluation

The empirical results comprehensively validate the advantages of integrating fuzzy structural priors with Bayesian inference in data-scarce environments. The primary metrics for evaluation were classification accuracy on the withheld test set and the Expected Calibration Error, which quantifies the discrepancy between the model's predicted confidence and its actual empirical accuracy. A lower Expected Calibration Error indicates that a model is better calibrated; it is confident when it is correct and uncertain when it is likely to be wrong, which is essential for safe deployment.

Table 2: Classification accuracy and Expected Calibration Error across different models under five percent supervision

Model Architecture	Accuracy (Medical)	ECE (Medical)	Accuracy (Industrial)	ECE (Industrial)
Deterministic MLP	68.4	0.45	71.2	0.280
Standard Neuro-Fuzzy	72.1	0.210	74.5	0.255
Bayesian Neural Network	70.8	0.125	73.9	0.110
Bayesian Fuzzy Network	78.5	0.065	81.3	0.058

As detailed in the experimental results, the deterministic Multilayer Perceptron exhibited the poorest performance across all benchmarks. Constrained by only five percent of the training data, the standard neural network severely overfitted, resulting in low generalization accuracy and highly inflated Expected Calibration Errors. The network produced extreme confidence scores even for test samples situated far from

the training distribution. The standard Neuro-Fuzzy system demonstrated improved accuracy over the plain Multilayer Perceptron, highlighting the value of the fuzzy structural prior in limiting hypothesis space complexity. However, without a probabilistic framework, its calibration remained suboptimal, producing falsely confident predictions on ambiguous test cases. The standard Bayesian Neural Network successfully addressed the calibration issue. By leveraging Monte Carlo dropout to approximate weight uncertainty, it significantly reduced the Expected Calibration Error compared to the deterministic models. However, its overall classification accuracy remained comparable to or slightly worse than the standard Neuro-Fuzzy model. This suggests that while the Bayesian Multilayer Perceptron knew when it was uncertain, its unconstrained architectural flexibility made it difficult to learn the correct decision boundaries from such a minimal amount of data. The proposed Bayesian Fuzzy Neural Network achieved the highest performance across all metrics. By embedding variational inference within the localized structure of the fuzzy rule base, the model achieved a remarkable classification accuracy of 78.5 on the medical dataset and 81.3 on the industrial dataset, significantly outperforming all baselines. More importantly, it achieved an exceptionally low Expected Calibration Error. The structural constraint provided by the fuzzy membership functions allowed the Bayesian inference process to converge more rapidly and find optimal posterior distributions even with minimal data. When evaluating out-of-distribution samples or samples lying on the complex decision boundaries of the synthetic dataset, the Bayesian Fuzzy Neural Network produced broad predictive distributions with high variance. The localized nature of the rules ensured that regions without training data retained their initial broad prior distributions, effectively translating a lack of data directly into high epistemic uncertainty. This dual capability of achieving superior generalization while maintaining rigorous and reliable confidence estimates confirms the profound theoretical and practical utility of the proposed hybrid architecture.

5. Conclusion

This paper has presented the Bayesian Fuzzy Neural Network, a novel and sophisticated architecture designed to address the critical challenges of uncertainty quantification and robust classification in environments characterized by limited supervision. By fundamentally re-engineering the traditional neuro-fuzzy inference system to incorporate probabilistic parameters and rigorous variational inference, the proposed framework successfully bridges the divide between structural interpretability and Bayesian uncertainty modeling. The methodological core of this research demonstrated that treating the parameters of fuzzy membership functions and rule consequents as probability distributions allows for the seamless propagation of both aleatoric and epistemic uncertainty throughout the decision-making process. The empirical evaluations conducted across synthetic, medical, and industrial datasets unequivocally highlight the superiority of the Bayesian Fuzzy approach when learning from highly constrained data volumes. Conventional deterministic models, and even purely structural fuzzy models, demonstrated significant vulnerabilities to overfitting and severe overconfidence. While standard Bayesian neural networks improved calibration, they struggled to extract highly accurate representations from minimal data due to their unconstrained topologies. In contrast, the proposed architecture leveraged the fuzzy partitioning of the feature space as an immensely powerful structural prior. This localized processing constrained the Bayesian optimization landscape, enabling the model to achieve state-of-the-art accuracy under a five percent supervision regime while simultaneously maintaining an exceptionally low Expected Calibration Error. The implications of this research are highly significant for the deployment of artificial

intelligence in safety-critical domains where data acquisition is inherently limited. The ability of a model to explicitly communicate its ignorance regarding specific regions of the input space enables human operators to intervene, triggers fail-safe mechanisms, and directs active learning strategies to acquire the most informative labels. While the variational inference process introduces additional computational complexity during the training phase compared to deterministic models, the resulting gains in reliability and safety heavily outweigh these overhead costs. Future research directions will focus on scaling this architecture to accommodate unstructured data modalities, such as images and sequential text, by integrating probabilistic fuzzy layers with convolutional and recurrent feature extractors. Ultimately, the fusion of fuzzy logic and Bayesian inference provides a highly promising, mathematically grounded pathway toward the development of trustworthy, resilient, and transparent machine learning systems.

References

- Hu, Z., Hu, M., & Li, Z. (2026). DynSupplyNet: A Dynamic Graph Neural Network with Temporal Fusion for Supply Chain Risk Prediction and Propagation Analysis. *Journal of Computer Science and Frontier Technologies*, 3(2), 1-13.
- Yang, Tianyue, and Xiao Xue. "Meno: Meanflow-enhanced neural operators for dynamical systems." *arXiv preprint arXiv:2604.06881* (2026).
- Wang, Jiacheng, et al. "A Dynamic Factor Gating Architecture with Market Regime Awareness for Stock Return Forecasting." (2026).
- Ma, Y., Qu, D., & Wang, Y. (2026). Dynamic community detection using class preserving time series generation with Fourier Markov diffusion. *Scientific Reports*. (<https://doi.org/10.1038/s41598-026-37699-1>)
- Qu, D., & Ma, Y. (2026). F2-CommNet: Fourier–Fractional neural networks with Lyapunov stability guarantees for hallucination-resistant community detection. *Frontiers in Computational Neuroscience*, 19, 1731452. (<https://doi.org/10.3389/fncom.2025.1731452>)
- Chen, Y., Lyu, N., Lang, S., Yan, H., Tao, Z., Ding, X., & Zhu, X. EconAI: Dynamic Persona Evolution and Memory-Aware Agents inEvolving Economic Environments. In *Workshop on Multi-Agent Learning and Its Opportunities in the Era of Generative AI*.
- Hu, L., & Shen, Y. (2026). A predictive analytics approach for forecasting global stock index returns using deep learning techniques. *Decision Analytics Journal*, 100685.
- Xue, Xiao, et al. "Fast-Forward Lattice Boltzmann: Learning Kinetic Behaviour with Physics-Informed Neural Operators." *arXiv preprint arXiv:2509.22411* (2025).
- Hao, Y., Xu, J., & Liao, S. (2026). RAPS-Net: A Risk-Aware CNN-LSTM Framework for Cross-Domain Risk Prediction and Dynamic Security Control in Cloud Payment Supply Chains. *Economics and Management Innovation*, 3(2), 18-27.
- Liu, J. (2025, May). Reinforcement learning-controlled subspace ensemble sampling for complex data structures. In *2025 IEEE 7th International Conference on Communications, Information System and Computer Engineering (CISCE)* (pp. 590-594). IEEE.
- Ho, S. M., & Liu, Y. (2023, June). Genie Breaks the Bottle: Ethics in Artificial Intelligence Adoption: ChatGPT; The Beginning and the End of Human Wisdom. In *Proceedings of the 2023 Computers and People Research Conference* (pp. 1-4).
- Lv, Qi, et al. "Spatial-temporal graph diffusion policy with kinematic modeling for bimanual robotic manipulation." *Proceedings of the Computer Vision and Pattern Recognition Conference*. 2025.

- Xu, Yuanchao, et al. "Reinforced data-driven estimation for spectral properties of Koopman semigroup in stochastic dynamical systems." arXiv preprint arXiv:2509.04265 (2025).
- Qu, D., Ma, Y., & Pyrozhenko, M. (2026). DISPEL-GNN: De-Illusion via Spectral Stability and Perturbation Bound-Enforced Learning for Community Detection with Risk-Aware Dynamic Attention in Graph Neural Networks. *Mathematics*, 14(4), 602. (<https://doi.org/10.3390/math14040602>)
- Wang, Z., Oh, S. K., Wang, Z., Fu, Z., Pedrycz, W., & Yoon, J. H. (2023). Design of progressive fuzzy polynomial neural networks through gated recurrent unit structure and correlation/probabilistic selection strategies. *Fuzzy Sets and Systems*, 470, 108656.
- Cheng, Xiaoyuan, et al. "How Does the Lagrangian Guide Safe Reinforcement Learning through Diffusion Models?." arXiv preprint arXiv:2602.02924 (2026).
- Guo, H., Lu, C., Bao, F., Pang, T., Yan, S., Du, C., & Li, C. (2023). Gaussian mixture solvers for diffusion models. *Advances in Neural Information Processing Systems*, 36, 25598-25626.
- Ma, Y., & Qu, D. (2025, November). GELNO-FD: gauge-equivariant Fourier liquid neural operators for interpretable Markovian Bayesian dynamics. In *Fifth International Conference on Advanced Algorithms and Signal Image Processing (AASIP 2025)* (Vol. 13967, pp. 177-183). SPIE.
- Yao, C., Zhan, Q., Cao, Z., Lin, Y., Li, D., Shao, Y., ... & Dong, Z. (2025). Generative ai for simulating real world dynamics applications and challenges.
- Dong, J., Koniusz, P., Chen, J., Xie, X., & Ong, Y. S. (2024). Adversarially robust few-shot learning via parameter co-distillation of similarity and class concept learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 28535-28544).
- Ma, Y., Qu, D., & Wang, Y. (2026). Tracking evolving communities in fake news cascades using temporal graphs. *Scientific Reports*. (<https://www.nature.com/articles/s41598-026-35175-4>)
- Ma, Y., & Qu, D. (2025, August). Geftnn-ba: A gauge-equivariant fourier transformer neural network with bayesian attention for trustworthy temporal dynamics. In *2025 2nd International Conference on Intelligent Perception and Pattern Recognition (IPPR)* (pp. 314-318). IEEE.
- Dong, J., Koniusz, P., Chen, J., & Ong, Y. S. (2024, September). Adversarially robust distillation by reducing the student-teacher variance gap. In *European Conference on Computer Vision* (pp. 92-111). Cham: Springer Nature Switzerland.
- Chen, L. (2026). Beyond external constraints: The missing dimension of ai governance. Available at SSRN 6449738.
- Xue, X., Wang, S., Yao, H. D., Davidson, L., & Coveney, P. V. (2024). Physics informed data-driven near-wall modelling for lattice Boltzmann simulation of high Reynolds number turbulent flows. *Communications Physics*, 7(1), 338.
- Xue, Xiao, et al. "Uni-Flow: a unified autoregressive-diffusion model for complex multiscale flows." arXiv preprint arXiv:2602.15592 (2026).
- Dong, J., Wang, Y., Lai, J. H., & Xie, X. (2022). Improving adversarially robust few-shot image classification with generalizable representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9025-9034).

- Ma, Y., Qu, D., & Pyrozhenko, M. (2026). Bio-RegNet: A Meta-Homeostatic Bayesian Neural Network Framework Integrating Treg-Inspired Immunoregulation and Autophagic Optimization for Adaptive Community Detection and Stable Intelligence. *Biomimetics*, 11(1), 48.
- Mnih, V., Badia, A. P., Mirza, M., Graves, A., Harley, T., Lillicrap, T., et al. (2016). Asynchronous methods for deep reinforcement learning. In *Proceedings of ICML*.
- Dong, J., Koniusz, P., Chen, J., Wang, Z. J., & Ong, Y. S. (2024). Robust distillation via untargeted and targeted intermediate adversarial samples. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 28432-28442).
- Zhong, Z., & Huang, Y. (2026). CRE-CNN-LSTM-XAI: An Explainable Deep Learning Framework for Risk Assessment of Real Estate Collateralized Financial Products. *Journal of Computer Science and Frontier Technologies*, 3(1), 127-140.
- Zhao, S., Qian, K., & Wen, C. (2022). Research on Used Car Transaction Cycle Based on Soft Voting. *Information Systems and Economics*, 3(1), 10-15.
- Li, W. W., Ziser, Y., Xie, Y., Cohen, S. B., & Ma, T. (2025, July). Tsprank: Bridging pairwise and listwise methods with a bilinear travelling salesman model. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1* (pp. 707-718).
- Qu, D., & Ma, Y. (2025). Magnet-bn: markov-guided Bayesian neural networks for calibrated long-horizon sequence forecasting and community tracking. *Mathematics*, 13(17), 2740.