

Article

# Stability-Enhanced Pseudo-Multiview Learning via Multiscale Grid Feature Extraction

Dat Ngo 

Department of Computer Engineering, Korea National University of Transportation,  
Chungju 27469, Republic of Korea; datngo@ut.ac.kr; Tel.: +82-43-841-5347

## Abstract

Pseudo-multiview learning improves classification by integrating complementary feature representations, but its performance degrades as the number of pseudo-views increases due to model collapse and ineffective feature scaling. This paper introduces a multiscale grid architecture that extracts structured, scale-adaptive features to stabilize evidence aggregation in pseudo-multiview learning. The proposed design enables efficient handling of difficult classification scenarios by enforcing balanced multiscale representation and reducing redundancy across pseudo-views. Extensive experiments on challenging real-world datasets, including BreakHis ( $40\times$ ,  $100\times$ ,  $200\times$ ,  $400\times$ ), Oxford-IIIT Pet, and Chest X-ray, demonstrate consistent gains in accuracy and stability over the original pseudo-multiview framework and other baseline models. The results confirm that grid-based multiscale feature extraction provides a reliable means to enhance pseudo-multiview learning, particularly in settings where prior methods struggled to generalize.

**Keywords:** image classification; multiview learning; multiscale feature representation

**MSC:** 68U10

## 1. Introduction

Image classification is a core problem in computer vision and serves as the foundation for numerous downstream applications. Early approaches depended on handcrafted descriptors such as scale-invariant feature transform (SIFT) [1], histogram of oriented gradients (HOG) [2], and color-based statistics, paired with classical classifiers including support vector machines (SVMs) [3] and  $k$ -nearest neighbors ( $k$ -NN) [4]. Although effective for constrained tasks, these pipelines require extensive feature engineering and often fail to capture complex visual structures.

Deep learning has reshaped image classification by enabling models to learn discriminative representations directly from raw images. Convolutional neural networks (CNNs) [5] and their large-scale variants such as EfficientNet [6], HorNet [7], ConvNeXt [8], and MaxVit [9] have achieved substantial gains on benchmarks like ImageNet [10]. These advances have established deep learning as the dominant framework across practical domains, including medical image analysis, autonomous systems, and surveillance, where high accuracy and generalization are essential.

Combining feature representations is a widely used strategy for improving classification performance. Feature concatenation [11], in particular, remains a simple yet effective method for aggregating complementary information extracted by deep models. However,



Academic Editor: Manuel G. Forero

Received: 26 January 2026

Revised: 20 March 2026

Accepted: 21 March 2026

Published: 23 March 2026

**Copyright:** © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

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

conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

this strategy can inflate model capacity and exacerbate the confidence calibration problem (CCP) [12], in which networks produce overconfident predictions that lack reliable uncertainty estimates. To mitigate calibration issues, uncertainty-aware classification methods have been explored through Bayesian [13,14] and non-Bayesian [15,16] formulations. Bayesian frameworks provide principled uncertainty quantification but incur high computational cost, while many non-Bayesian approaches support only single-view data, limiting their applicability.

Standard neural classifiers rely on a softmax output, which provides normalized scores but does not explicitly model uncertainty. Alternatively, the Dirichlet distribution represents a probability density over the simplex and can encode both class evidence and its associated uncertainty. Subjective logic [17], which builds on Dirichlet-based modeling, offers a principled means to express belief, uncertainty, and prior knowledge within the same analytical framework.

Pseudo-multiview learning [18] adopts subjective logic to integrate information from multiple views by modeling each view with a variational Dirichlet distribution and fusing them at the evidence level. This approach preserves uncertainty throughout the decision process and avoids the drawbacks of direct feature or probability concatenation. While effective, its computational burden increases as more pseudo-views are introduced, leading to model collapse in difficult classification scenarios.

To address the above limitation, this work introduces a multiscale grid architecture designed to extract structured, diverse features that stabilize pseudo-multiview evidence fusion. The grid representation reduces redundancy across pseudo-views and mitigates the collapse effects observed in prior frameworks. Extensive experiments on challenging real-world datasets—including BreakHis at multiple magnification levels, Oxford-IIIT Pet, and Chest X-ray—demonstrate substantial improvements in classification accuracy and stability.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 details the proposed multiscale grid-based pseudo-multiview learning method. Section 4 reports experimental evaluations, and Section 5 concludes the paper.

## 2. Related Work

### 2.1. Multiview Learning

Multiview learning [19] is a machine learning paradigm in which multiple representations of the same underlying data are jointly exploited to enhance model performance. Each view captures a distinct aspect of the data distribution, and integrating these complementary perspectives often leads to more reliable and expressive feature representations. This paradigm has been successfully applied across diverse domains such as computer vision, natural language processing, and bioinformatics, where heterogeneous information sources are common.

A number of studies have explored different strategies for leveraging multiview information. For example, Zhang et al. [20] introduce a multiview framework for Alzheimer's disease diagnosis that simultaneously models the relationships between features and class labels while capturing higher-order dependencies across multiple kernel matrices. In computer vision, new multiview models have been proposed to address scenarios where certain views are missing or incomplete [21], enabling flexible fusion of partial information. In the field of program analysis, Long et al. [22] develop a multiview graph-based representation method that combines graph neural networks with multiview embedding techniques to improve code understanding.

Multiview learning has also been actively studied in clustering applications. A series of works by Liu et al. [23–25] propose methods for integrating incomplete clustering matrices derived from partially observed views and address the challenge of selecting appropriate

hyperparameters for unsupervised multiview settings. Additional contributions in this direction [26,27] demonstrate that multiview information, when properly aggregated, can significantly improve clustering quality in complex multi-source environments.

## 2.2. Uncertainty-Aware and Pseudo-Multiview Learning

Reliable uncertainty estimation plays a critical role in safety-critical applications, where overconfident predictions may lead to incorrect decisions. Conventional deterministic neural networks provide point estimates through softmax outputs but do not explicitly quantify uncertainty [12]. To address this limitation, a variety of uncertainty-aware methods have been developed.

Sampling-based approaches such as Monte Carlo Dropout [28] treat stochastic forward passes as approximate Bayesian inference, offering a lightweight means of estimating predictive variance. Other methods incorporate uncertainty directly into the architecture—for example, uncertainty-aware attention mechanisms [29] modulate feature importance using confidence estimates, while evidential deep learning [30] models predictions as evidence for a Dirichlet distribution, enabling the network to simultaneously infer belief and associated uncertainty. Although effective, these approaches generally operate in single-view settings and do not exploit complementary information across multiple feature perspectives.

Pseudo-multiview learning (PML) [18] extends uncertainty-aware modeling to a multiview setting by generating multiple internal representations, or psuedo-views, from a single input. Instead of concatenating features or combining class-probability outputs, PML interprets non-negative evidence from each psuedo-view as parameters of a Dirichlet distribution. Subjective logic is then used to map this evidence into belief and uncertainty, and the final decision is obtained by fusing opinions through an averaging rule. This formulation enables the model to down-weight unreliable views while emphasizing those that provide stronger evidence.

Figure 1 shows the overall flow of PML, where the evidence for a given view is denoted by  $e_i$ . The corresponding Dirichlet parameter is computed as  $\alpha_i = 1 + e_i$ . The opinion  $\omega_i = \{b_i, u_i, a\}$  is then obtained using the following:

$$b_i = e_i/S_i, \quad (1)$$

$$u_i = K/S_i, \quad (2)$$

where  $S_i = \sum_k e_i(k)$  is the total evidence for the  $i$ -th view and  $K$  is the number of classes. Here,  $b_i$  denotes the belief,  $u_i$  the uncertainty, and  $a$  the base rate. In the absence of prior knowledge, a uniform base rate  $a = 1/K$  can be used. As psuedo-views are generated simultaneously, their opinions are fused using the averaging rule. To illustrate the fusion process (the last stage in Figure 1), Table 1 provides an example of fusing two opinions. The fused belief for the first class (0.52) lies between the original beliefs (0.6 and 0.4), and the fused uncertainty (0.24) also falls between the original uncertainties (0.2 and 0.3), demonstrating that averaging fusion yields a consistent and uncertainty-aware combination of the two inputs.

In the broader multiview learning literature, the idea of artificial view generation has been actively explored to compensate for missing views or enrich representation diversity. Contrastive learning methods such as SimCLR [31] create multiple augmented views of the same sample to strengthen representation invariance. GAN-based view synthesis frameworks [32] generate synthetic viewpoints to mimic unavailable camera angles or modalities. Rotation-based artificial view augmentation has been proposed to improve multiview embedding robustness [33]. In clustering, synthetic feature generation is employed to emulate incomplete or corrupted views [34]. These studies collectively highlight

the benefits of leveraging constructed or transformed views when genuine multiview observations are limited or partially missing.

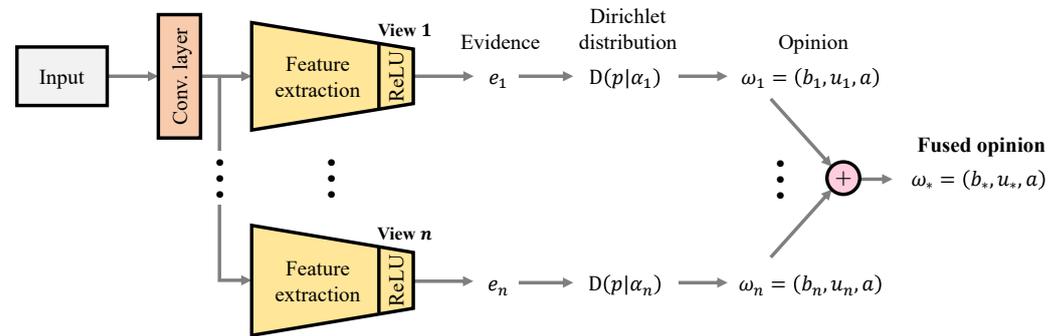


Figure 1. Illustration of pseudo-multiview learning.

Table 1. Summary of the opinion fusion process. A binary classification problem with two opinions is considered.

Step	Formula	Example
Input	$\omega_1 = \{b_1, u_1, a\}$ $\omega_2 = \{b_2, u_2, a\}$	$b_1 = \{0.6, 0.2\}, u_1 = 0.2, a = \{0.5, 0.5\}$ $b_2 = \{0.4, 0.3\}, u_2 = 0.3$
Evidence conversion	$r_1(k) = W \cdot b_1(k) / u_1$ $r_2(k) = W \cdot b_2(k) / u_2$ W is a non-informative weight	$r_1 = \{6, 2\}$ $r_2 = \{2.67, 2\}$ $W = 2$
Opinion fusion	$r_*(k) = [r_1(k) + r_2(k)] / 2$ $S_* = W + \sum_k r_*(k)$ $b_*(k) = r_*(k) / S_*$ $u_*(k) = W / S_*$	$r_* = \{4.33, 2\}$ $S_* = 8.33$ $b_* = \{0.52, 0.24\}$ $u_* = 0.24$
Fused opinion	$\omega_* = \{b_*, u_*, a\}$	$b_* = \{0.52, 0.24\}, u_* = 0.24,$ $a = \{0.5, 0.5\}$

Despite its advantages, conventional PML faces a scalability challenge. As the number of psuedo-views increases, the dimensionality of the evidence space grows rapidly, resulting in large parameter counts and unstable optimization. Fully independent full-resolution branches exacerbate this issue, often leading to convergence failures or degraded performance on complex datasets, as demonstrated in Appendix A. Furthermore, conventional PML does not include cross-branch information flow, limiting its ability to exploit shared structure across views.

Grid-structured neural architectures have recently emerged as an effective mechanism for organizing feature extraction across multiple spatial or semantic scales. Unlike traditional sequential or parallel networks, grid-based models introduce structured lateral and vertical connections that enable information to flow across scales while maintaining computational efficiency. Such representations have been shown to facilitate deeper supervision, improve gradient propagation, and enhance feature reuse. Examples include MIRNet for image restoration [35], MGFNet for multimodal semantic segmentation [36], and hierarchical multiscale feature fusion networks for medical image classification [37]. These approaches demonstrate that grid-based connectivity can effectively encode hierarchical relationships between feature maps, reduce redundancy, and stabilize optimization in deep architectures.

These findings motivate the development of the proposed multiscale grid architecture. By introducing structured interactions between psuedo-views and generating additional views through multiscale downsampling rather than full-resolution replication, the pro-

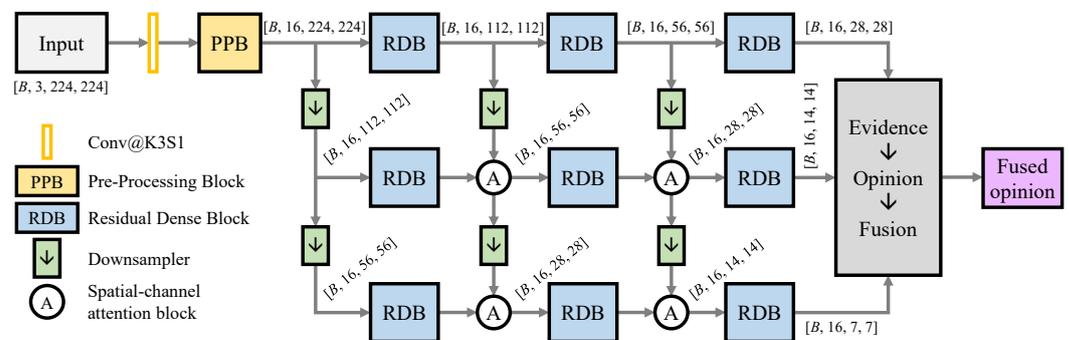
posed approach reduces parameter growth, stabilizes training, and provides a more reliable foundation for uncertainty-aware multiview learning.

### 3. Proposed Method

Figure 2 illustrates the overall architecture (inspired by GridDehazeNet+ [38]) for three psuedo-views. The model first extracts base-level features from the input using a Pre-Processing Block (PPB) and then propagates these features through a grid of Residual Dense Blocks (RDBs). Downsampling operators connect vertical branches, forming lower-resolution psuedo-views, while lateral skip connections maintain information flow within each resolution level. Each branch is followed by a Spatial-Channel Attention Block (SCAB) to enhance discriminative feature learning. For each psuedo-view, the final features are transformed into non-negative evidence vectors, which are mapped to Dirichlet opinions and fused using subjective logic.

Let  $[B, C, H, W]$  denote the batch size, channel number, height, and width of the feature maps in each psuedo-view. Given the input of  $[B, 3, 224, 224]$ , the PPB produces an initial feature map of size  $[B, 16, 224, 224]$  for the first row, which processes features at full resolution. Meanwhile, the second and third rows apply downsampling to create psuedo-views at  $[B, 16, 112, 112]$  and  $[B, 16, 56, 56]$ , respectively. Each RDB refines features through dense connections, and the SCAB applies attention mechanisms to enhance feature quality. The size of each feature map is reduced as it passes through the grid, with the final outputs for the three psuedo-views being  $[B, 16, 28, 28]$ ,  $[B, 16, 14, 14]$ , and  $[B, 16, 7, 7]$ . These features are then transformed into evidence vectors for classification, as discussed earlier in Section 2.2.

The multiscale grid architecture enables the model to capture complementary information across different resolutions while controlling parameter growth, thus addressing the limitations of the conventional PML. The following subsections provide detailed descriptions of each component in the architecture.



**Figure 2.** Overall architectural design of the proposed stability-enhanced pseudo-multiview learning method for three psuedo-views. Conv@K3S1 denotes a convolutional layer with kernel size  $3 \times 3$  and stride 1. The notation  $[B, C, H, W]$  denotes the batch size, channel number, height, and width.

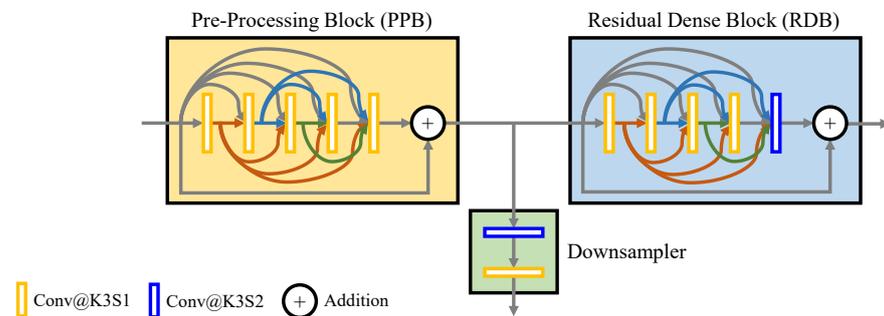
#### 3.1. Multiscale Grid Architecture

The multiscale grid consists of three rows corresponding to three psuedo-views and three columns representing progressive feature refinement stages. The top row receives full-resolution features from the PPB. Subsequent rows receive progressively downscaled inputs via learnable downsamplers. This grid architecture offers three advantages:

- It creates a structured set of psuedo-views with controlled complexity.
- It reduces the number of parameters per view, mitigating the model collapse problem observed in conventional PML.

- It encourages multiscale feature complementarity that is beneficial for difficult classification tasks.

Each node in the grid is an RDB, while the edges correspond to downsampling connections. Details of these modules are shown in Figure 3.



**Figure 3.** Architectural designs of pre-processing, residual dense, and downsampler blocks. Connections are shown in different colors for ease of viewing.

### 3.2. Pre-Processing Block

The PPB (Figure 3, left) extracts shallow features from the input image and prepares them for multiscale processing. It consists of a sequence of convolutional layers with kernel size  $3 \times 3$ , each followed by a ReLU activation. A final residual addition stabilizes gradient propagation and preserves low-frequency image structure. The PPB produces the initial feature map for the top-left node of the grid.

### 3.3. Residual Dense Block

Each RDB (Figure 3, right) refines features through a series of densely connected convolutional layers. Features from all preceding layers within the block are concatenated before being passed to the next layer, enabling effective information reuse and improving feature diversity. A local residual connection is applied from the block input to its output to alleviate vanishing gradients and support deeper feature extraction. RDBs form the fundamental processing units of the grid.

### 3.4. Downsampling for Multiscale View Generation

Downsampling modules connect vertically adjacent nodes in the grid (Figure 2). Each downsampler applies a convolution-based reduction in spatial resolution, creating additional psuedo-views with substantially fewer parameters than full-resolution branches. By generating views from downscaled features, this strategy enables the use of more psuedo-views without causing memory overflow or convergence failure, directly addressing the limitations described in Section 2.2.

### 3.5. Spatial-Channel Attention Block

To enhance discriminative feature representations, each branch passes through a Spatial-Channel Attention Block (SCAB), as illustrated in Figure 4. The SCAB integrates the following elements.

- The Channel Attention Block (CAB) applies global max and average pooling to estimate channel-wise importance. A shared Multi-Layer Perceptron (MLP) infers channel attention weights, which are applied via pixel-wise multiplication.
- The Spatial Attention Block (SAB) aggregates channel information via spatial max and average pooling, followed by a convolutional filter to compute spatial attention weights, which are also applied via pixel-wise multiplication.

The SCAB performs channel and spatial attention to emphasize informative regions and suppress background information, thus improving the quality of features fed into the next RDB.

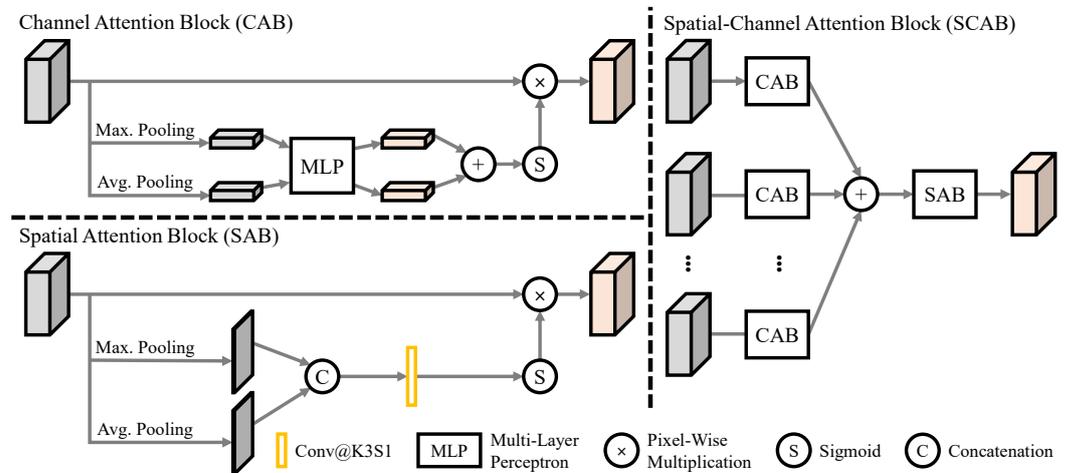


Figure 4. Architectural design of the Spatial-Channel Attention Block.

### 3.6. Evidence Extraction and Opinion Fusion

For each psuedo-view, the final feature tensor is transformed by a  $1 \times 1$  convolution layer with ReLU activation to produce a non-negative evidence vector  $e_i$ . The corresponding Dirichlet parameters are computed as  $\alpha_i = e_i + 1$ . The Dirichlet strength  $S_i = \sum_k \alpha_i(k)$  yields the belief and uncertainty components of the subjective logic opinion, detailed as follows.

- Belief:  $b_i(k) = e_i(k) / S_i$
- Uncertainty:  $u_i = K / S_i$
- Base rate:  $a = 1 / K$  (uniform)

Opinions from all psuedo-views are fused using the averaging rule, which yields a balanced and uncertainty-aware aggregated opinion. The maximum-belief class of the fused opinion is taken as the final prediction.

The proposed multiscale grid architecture provides a principled solution to the scalability limitations of the conventional PML. By forming psuedo-views through downscaled feature branches and enforcing structured information flow via the grid topology, the model avoids parameter explosion and training instability. The combination of PPB, RDB, down-sampling, and SCAB enhances multiscale feature extraction while preserving subjective logic-based uncertainty modeling.

### 3.7. Loss Function

Let  $x_i$  denote the input image and  $y_i$  the corresponding class label. The model generates  $N$  psuedo-views, each producing an evidence vector  $e_i$  that is at the output of the last RDB in each row. The evidence vectors are transformed into Dirichlet parameters  $\alpha_i = e_i + 1$ . The cross-entropy loss  $\mathcal{L}_{CE}$  is applied with an additional KL-divergence regularization term  $\mathcal{L}_{KL}$  to encourage the model to produce well-calibrated uncertainty estimates. The overall loss function is defined as follows:

$$\mathcal{L}(x_i, y_i) = \sum_{j=1}^N \left( \mathcal{L}_{CE}^j(x_i, y_i) + \lambda \mathcal{L}_{KL}^j(x_i, y_i) \right), \tag{3}$$

where  $\lambda$  is a hyperparameter that balances the two loss components. For the  $j$ -th psuedo-view, the cross-entropy and KL-divergence losses are computed as follows:

$$\mathcal{L}_{\text{CE}}(x_i, y_i) = \sum_{k=1}^K y_i(k) (\psi(S_i) - \psi(\alpha_i(k))), \quad (4)$$

$$\mathcal{L}_{\text{KL}}(x_i, y_i) = \log \frac{\Gamma(\theta_i)}{\Gamma(K) \prod_{k=1}^K \Gamma(\alpha_i(k))} + \sum_{k=1}^K (\alpha_i(k) - 1) (\psi(\alpha_i(k)) - \psi(\theta_i)), \quad (5)$$

where  $\theta_i = \sum_{k=1}^K \alpha_i(k)$  is the total Dirichlet strength,  $\Gamma(\cdot)$  is the gamma function, and  $\psi(\cdot)$  is the digamma function.  $S_i$  and  $K$  are defined earlier in Section 2.2 as the total evidence and the number of classes, respectively.

## 4. Experimental Results

### 4.1. Experimental Settings

The proposed multiscale grid architecture is evaluated on three real-world datasets drawn from distinct application domains: microscopic pathology imaging (BreakHis [39]), natural image categorization (Oxford-IIIT Pet [40]), and medical radiography (Chest X-ray [41]). Training and testing the model separately on multiple datasets from distinct domains is a commonly accepted way to evaluate the architectural generalizability of the method [42,43]. It demonstrates that the proposed architecture adapts well to diverse visual environments, even though it does not measure out-of-distribution performance. It is also worth noting that the conventional PML exhibits limited performance on these datasets, which further underscores the need for a more robust architecture.

The BreakHis dataset comprises 9109 breast tumor microscopy images collected from 82 patients. Images are provided at four magnification levels ( $40\times$ ,  $100\times$ ,  $200\times$ , and  $400\times$ ) and are categorized into benign and malignant classes, offering a diverse set of examples for binary classification under varying visual scales.

The Oxford-IIIT Pet dataset contains 37 fine-grained categories of cats and dogs, with roughly 200 images per category. The dataset presents substantial intra-class variability due to changes in pose, illumination, and appearance, making it a challenging benchmark for evaluating feature discrimination ability.

The Chest X-ray dataset consists of 5863 pediatric radiographs labeled as either pneumonia or normal. Images were acquired as part of routine clinical practice at Guangzhou Women and Children's Medical Center. Low-quality scans were removed, and the annotations were verified by multiple radiologists to ensure reliable diagnostic labels.

A five-fold cross-validation protocol is used in all evaluations. Hyperparameters are selected using the Microsoft Neural Network Intelligence toolkit [44]. The learning rate is searched within the interval  $[10^{-4}, 10^{-1}]$ , and batch sizes are drawn from  $\{8, 32, 64, 256, 512, 1024\}$ . Training is terminated early if no improvement is observed for ten consecutive epochs.

### 4.2. Comparison with Conventional Pseudo-Multiview Learning

#### 4.2.1. Performance Comparison

To provide a direct comparison with the proposed model, the architecture illustrated in Figure 5 is implemented. In this baseline, each branch operates independently, and no cross-branch interactions are applied. All input images are resized to  $224 \times 224$ , allowing experiments with two, three, and four psuedo-views.

Table 2 reports the AUC and ACC values for both the conventional PML and the proposed model across all datasets. Several consistent patterns emerge. First, conventional PML exhibits modest improvement when increasing the number of views from two to

three, as additional branches introduce complementary feature perspectives. However, when extended to four full-resolution branches, performance deteriorates due to the rapid growth in model parameters. This behavior aligns with the known limitations of PML. Its independent parallel branches lead to unstable optimization, slow convergence, and degraded classification accuracy when the model becomes excessively large.

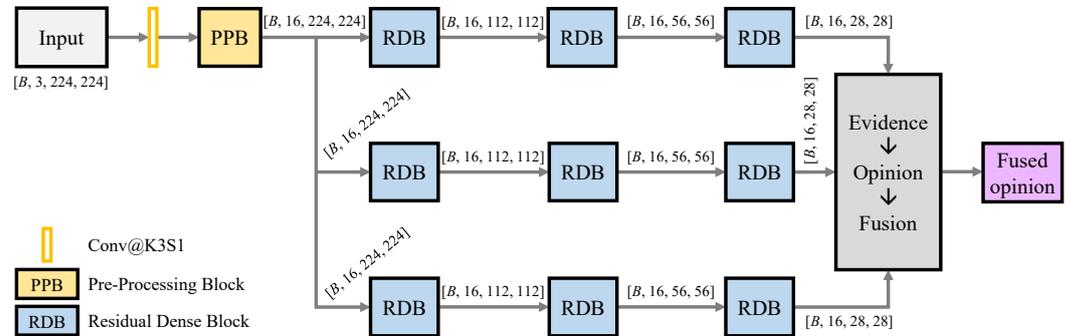


Figure 5. Overall architectural design of the baseline model (the conventional PML) for three psuedo-views.

Table 2. Performance comparison between conventional pseudo-multiview learning and the proposed method. The best results are in bold.

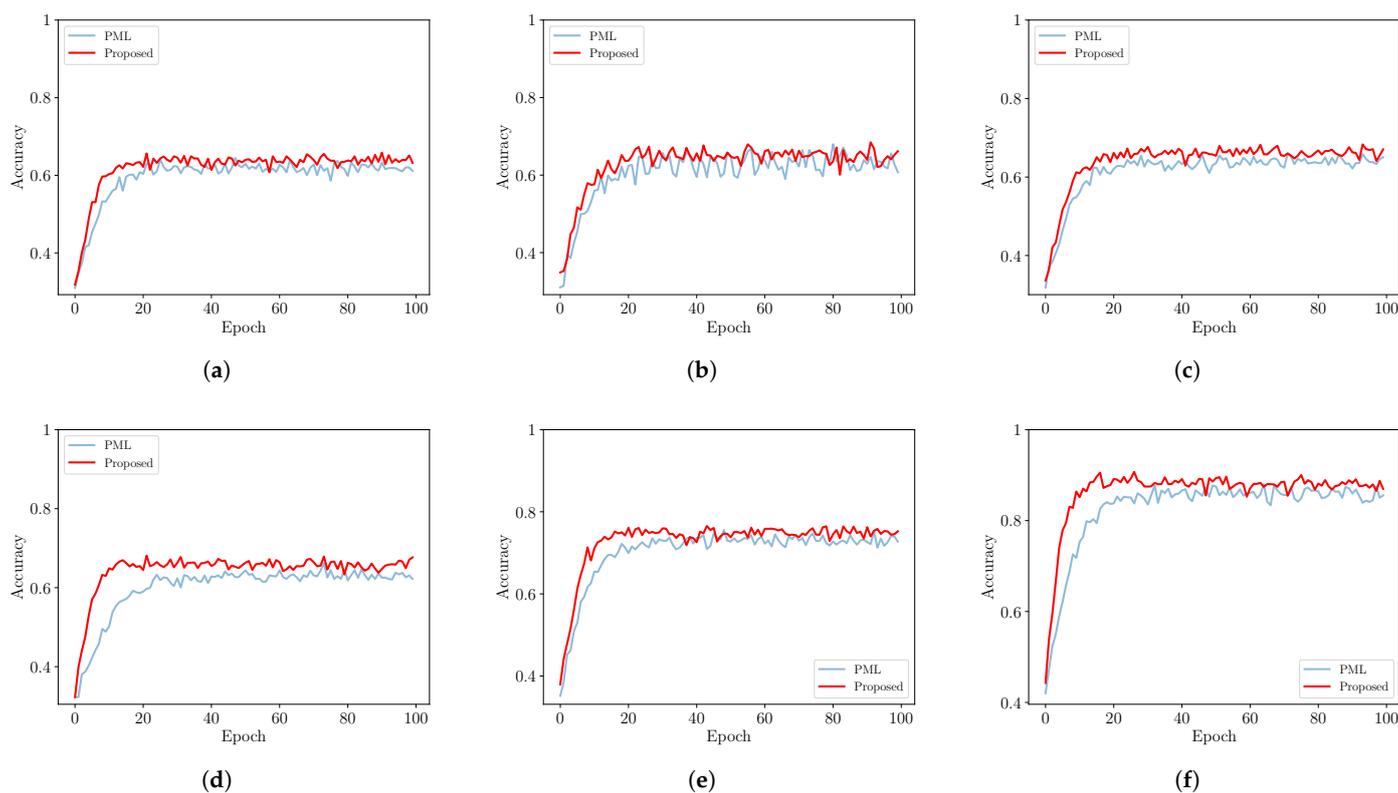
Dataset	Number of Views	AUC		ACC	
		Conventional	Proposed	Conventional	Proposed
BreakHis 40×	2	0.590 ± 0.039	<b>0.618 ± 0.037</b>	0.622 ± 0.040	<b>0.645 ± 0.038</b>
	3	0.611 ± 0.041	<b>0.638 ± 0.037</b>	0.634 ± 0.042	<b>0.662 ± 0.036</b>
	4	0.565 ± 0.048	<b>0.654 ± 0.035</b>	0.598 ± 0.049	<b>0.676 ± 0.034</b>
BreakHis 100×	2	0.652 ± 0.047	<b>0.672 ± 0.035</b>	0.631 ± 0.072	<b>0.651 ± 0.058</b>
	3	0.669 ± 0.044	<b>0.685 ± 0.033</b>	0.645 ± 0.067	<b>0.668 ± 0.054</b>
	4	0.621 ± 0.051	<b>0.701 ± 0.032</b>	0.602 ± 0.073	<b>0.683 ± 0.049</b>
BreakHis 200×	2	0.688 ± 0.027	<b>0.707 ± 0.031</b>	0.641 ± 0.053	<b>0.668 ± 0.045</b>
	3	0.703 ± 0.025	<b>0.723 ± 0.028</b>	0.654 ± 0.049	<b>0.684 ± 0.042</b>
	4	0.656 ± 0.036	<b>0.739 ± 0.026</b>	0.612 ± 0.057	<b>0.701 ± 0.040</b>
BreakHis 400×	2	0.690 ± 0.032	<b>0.711 ± 0.029</b>	0.635 ± 0.029	<b>0.662 ± 0.027</b>
	3	0.702 ± 0.030	<b>0.726 ± 0.027</b>	0.648 ± 0.031	<b>0.676 ± 0.026</b>
	4	0.655 ± 0.038	<b>0.741 ± 0.025</b>	0.603 ± 0.038	<b>0.694 ± 0.023</b>
Oxford-IIIT Pet	2	0.818 ± 0.022	<b>0.832 ± 0.028</b>	0.733 ± 0.029	<b>0.751 ± 0.030</b>
	3	0.829 ± 0.020	<b>0.845 ± 0.026</b>	0.744 ± 0.027	<b>0.764 ± 0.030</b>
	4	0.688 ± 0.047	<b>0.861 ± 0.025</b>	0.609 ± 0.033	<b>0.778 ± 0.027</b>
Chest X-ray	2	0.917 ± 0.010	<b>0.930 ± 0.012</b>	0.869 ± 0.012	<b>0.881 ± 0.013</b>
	3	0.924 ± 0.009	<b>0.938 ± 0.010</b>	0.875 ± 0.011	<b>0.892 ± 0.010</b>
	4	0.683 ± 0.015	<b>0.945 ± 0.018</b>	0.599 ± 0.017	<b>0.904 ± 0.012</b>

In contrast, the proposed multiscale PML consistently demonstrates monotonic performance gains when increasing the number of psuedo-views. The architectural elements of the model-downsampled branches, the grid-structured topology, and the Spatial-Channel Attention Modules allow new views to be introduced without inflating the parameter count. As a result, the proposed method achieves superior accuracy and AUC across all view counts and datasets.

In addition, Figure 6 provides insight into the convergence behavior of both models. Across all datasets, the proposed method attains higher accuracy earlier in training and stabilizes more rapidly than conventional PML. For example, on BreakHis 40× and 100×, the proposed model reaches near-optimal performance within the first 20~30 epochs,

whereas the conventional PML converges more slowly and exhibits larger fluctuations. Similar trends appear in the Oxford-IIIT Pet and Chest X-ray datasets, where the proposed method consistently demonstrates faster stabilization and higher accuracy at intermediate epochs. These convergence plots highlight the benefits of multiscale feature integration and cross-branch interactions, both of which contribute to a smoother optimization landscape and more efficient training dynamics.

Overall, the empirical results confirm that the proposed multiscale grid architecture not only overcomes the scalability limitations of conventional PML but also accelerates convergence and enhances predictive performance. This improvement enables reliable and efficient multiview learning even when the number of psuedo-views increases or the dataset exhibits substantial intra-class variability.



**Figure 6.** Model convergence on different real-world datasets. (a) BreakHis 40×; (b) BreakHis 100×; (c) BreakHis 200×; (d) BreakHis 400×; (e) Oxford-IIIT Pet; (f) Chest X-ray.

#### 4.2.2. Parameter Count Comparison

To quantify the computational advantages of the proposed multiscale grid architecture, a comparison of the number of trainable parameters against the conventional PML model under different numbers of psuedo-views is conducted. Table 3 summarizes the parameter counts for both approaches when configured with two, three, and four views.

**Table 3.** Parameter count comparison between the conventional PML and the proposed method.

Number of Views	PML	Proposed	Reduction Rate
2	269,812	266,042	1.40%
3	398,822	386,578	3.07%
4	527,832	505,834	4.17%

The results show that the proposed method consistently requires fewer parameters than conventional PML. With two psuedo-views, the reduction is modest (1.40%), reflecting

the structural similarity between the two-branch configurations. However, as the number of views increases, the benefit becomes more pronounced. At three and four views, the proposed architecture achieves parameter reductions of 3.07% and 4.17%, respectively. In conventional PML, each additional psuedo-view introduces a full-resolution branch with no cross-scale connections, leading to near-linear parameter growth. In contrast, the proposed method constructs additional psuedo-views through multiscale downsampling within a grid topology, allowing lower-resolution branches to reuse upstream features and avoid duplicating full feature extractors.

These reductions, while moderate in absolute magnitude, have two important implications. First, they directly contribute to the improved training stability by preventing parameter explosion as the number of views increases. Second, they demonstrate that the multiscale grid architecture provides a more efficient representation strategy, enabling the integration of multiple psuedo-views without incurring the prohibitive computational cost associated with conventional PML.

Overall, the comparison confirms that the proposed method offers a more scalable and computationally efficient framework for pseudo-multiview learning, particularly when the number of views is increased.

### 4.3. Comparison with Uncertainty-Aware Methods

To further evaluate the effectiveness of the proposed multiscale PML framework, a comparison against three widely used uncertainty-aware classification methods—Monte Carlo Dropout (MCDO) [28], uncertainty-aware attention (UA) [29] and evidential deep learning (EDL) [30]—is conducted. These methods serve as strong baselines in uncertainty modeling and have been applied extensively in safety-critical applications. All three approaches operate in a single-view setting and are implemented on top of the first branch of the architecture in Figure 2 to ensure a fair comparison. By contrast, both the conventional PML and the proposed method use two psuedo-views.

MCDO estimates predictive uncertainty by performing multiple stochastic forward passes with dropout activated at inference time, providing a lightweight approximation to Bayesian inference. UA incorporates uncertainty information directly into the attention mechanism to highlight reliable features and suppress ambiguous regions. EDL models predictions as evidence for a Dirichlet distribution, enabling simultaneous estimation of belief and uncertainty without relying on sampling. Together, these methods represent the dominant strategies for uncertainty-aware deep learning.

Table 4 summarizes the AUC and ACC values achieved by all approaches across the BreakHis, Oxford-IIIT Pet, and Chest X-ray datasets.

**Table 4.** Performance comparison between uncertainty-aware approaches and the proposed method. The best results are in bold.

Metric	Dataset	MCDO	UA	EDL	PML	Proposed	
AUC	BreakHis	40×	0.567 ± 0.033	0.582 ± 0.029	0.585 ± 0.028	0.590 ± 0.039	<b>0.618 ± 0.037</b>
		100×	0.604 ± 0.030	0.626 ± 0.028	0.637 ± 0.026	0.652 ± 0.047	<b>0.672 ± 0.035</b>
		200×	0.671 ± 0.027	0.687 ± 0.023	0.694 ± 0.021	0.688 ± 0.027	<b>0.707 ± 0.031</b>
		400×	0.650 ± 0.034	0.671 ± 0.032	0.683 ± 0.030	0.690 ± 0.032	<b>0.711 ± 0.029</b>
	Oxford-IIIT Pet	0.781 ± 0.025	0.801 ± 0.020	0.808 ± 0.019	0.818 ± 0.022	<b>0.832 ± 0.028</b>	
	Chest X-ray	0.891 ± 0.011	0.905 ± 0.009	0.913 ± 0.009	0.917 ± 0.010	<b>0.930 ± 0.012</b>	
ACC	BreakHis	40×	0.586 ± 0.027	0.605 ± 0.024	0.616 ± 0.022	0.622 ± 0.040	<b>0.645 ± 0.038</b>
		100×	0.597 ± 0.026	0.612 ± 0.024	0.625 ± 0.025	0.631 ± 0.072	<b>0.651 ± 0.058</b>
		200×	0.602 ± 0.025	0.620 ± 0.022	0.631 ± 0.021	0.641 ± 0.053	<b>0.668 ± 0.045</b>
		400×	0.599 ± 0.028	0.616 ± 0.025	0.629 ± 0.023	0.635 ± 0.029	<b>0.662 ± 0.027</b>
	Oxford-IIIT Pet	0.688 ± 0.025	0.712 ± 0.020	0.720 ± 0.018	0.733 ± 0.029	<b>0.751 ± 0.030</b>	
	Chest X-ray	0.833 ± 0.015	0.850 ± 0.014	0.861 ± 0.013	0.869 ± 0.012	<b>0.881 ± 0.013</b>	

Several consistent observations emerge from Table 4. First, among the three single-view uncertainty-aware methods, EDL generally delivers the strongest overall performance, followed by UA and then MCDO. This observation aligns with the expected behavior. EDL benefits from explicit evidence modeling, while MCDO relies on sampling-based approximations that tend to produce higher variance.

Second, conventional PML achieves competitive results relative to these baselines, particularly on datasets with moderate complexity. However, as it uses two independent full-resolution branches, its representational capacity is limited by parameter growth and the lack of cross-branch interactions.

Most importantly, the proposed multiscale PML consistently outperforms all benchmark methods across every dataset and evaluation metric. The improvement is particularly notable on the BreakHis dataset, where multiscale visual characteristics are critical for accurate prediction. For example, at  $40\times$  magnification, the proposed method raises AUC from 0.590 (PML) and 0.567 (MCDO) to 0.618, while similar gains are observed in accuracy. Substantial improvements are also seen at higher magnifications, where conventional methods tend to struggle due to the variability of fine-grained tissue structures.

On the Oxford-IIIT Pet dataset, the proposed model reaches an AUC of 0.832, outperforming EDL (0.808) and UA (0.801). The advantage persists on the Chest X-ray dataset, where the proposed approach achieves the highest AUC (0.930) and ACC (0.881), demonstrating its superiority over benchmark methods.

These results highlight two key strengths of the proposed approach. First, the multiscale grid architecture enables the extraction of complementary features at different resolutions, improving the expressiveness of the model relative to single-view baselines. Second, subjective-logic-based fusion provides a stable mechanism for combining multiscale evidence while preserving uncertainty, allowing the model to down-weight unreliable psuedo-views and enhance overall prediction reliability.

Overall, the proposed method achieves the best performance among all evaluated uncertainty-aware techniques, confirming the advantage of combining multiscale feature extraction with principled opinion fusion under subjective logic.

#### 4.4. Ablation Study

##### 4.4.1. Multiscale Grid Architecture

To analyze the contribution of individual components within the proposed multiscale grid architecture, an ablation study is conducted, focusing on two key modules: the downsampler and the Spatial-Channel Attention Block (SCAB). Four configurations are evaluated:

- the baseline model without downsampler and SCAB;
- downsampler only (SCABs are replaced with addition operations);
- SCAB only (downsamplers are replaced with identity operations);
- the full model incorporating both modules.

The results are summarized in Table 5. Across all datasets, the baseline configuration exhibits the lowest AUC and ACC values, reflecting the limitations of conventional PML where each full-resolution branch operates independently and lacks structured multiscale interactions. Introducing the downsampler alone yields only minor improvements. This behavior is expected, as multiscale features provide additional complementary information, but without SCAB the model cannot fully exploit cross-scale dependencies or emphasize salient regions. The incremental improvements observed in this setting confirm that downsampling by itself is insufficient to substantially enhance feature quality.

In contrast, enabling SCAB alone leads to a clear performance gain on every dataset. SCAB's spatial and channel attention mechanisms strengthen informative features and

suppress background noise, allowing the model to generate higher-quality features. The improvements in both AUC and ACC demonstrate that attention plays a central role in enhancing the discriminative power of psuedo-views.

The full configuration, which combines downsampling and SCAB within the grid, consistently achieves the highest performance. This result confirms that SCAB benefits significantly from the multiscale feature hierarchy created by the downsampler. The grid arrangement allows information to flow across resolution levels, enabling SCAB to operate on richer and more diverse feature sets. As a result, the full model attains superior accuracy and stability across all datasets, validating the design of the proposed architecture.

**Table 5.** Effect of downsampler and SCAB on AUC and ACC across all datasets. The best results are in bold.

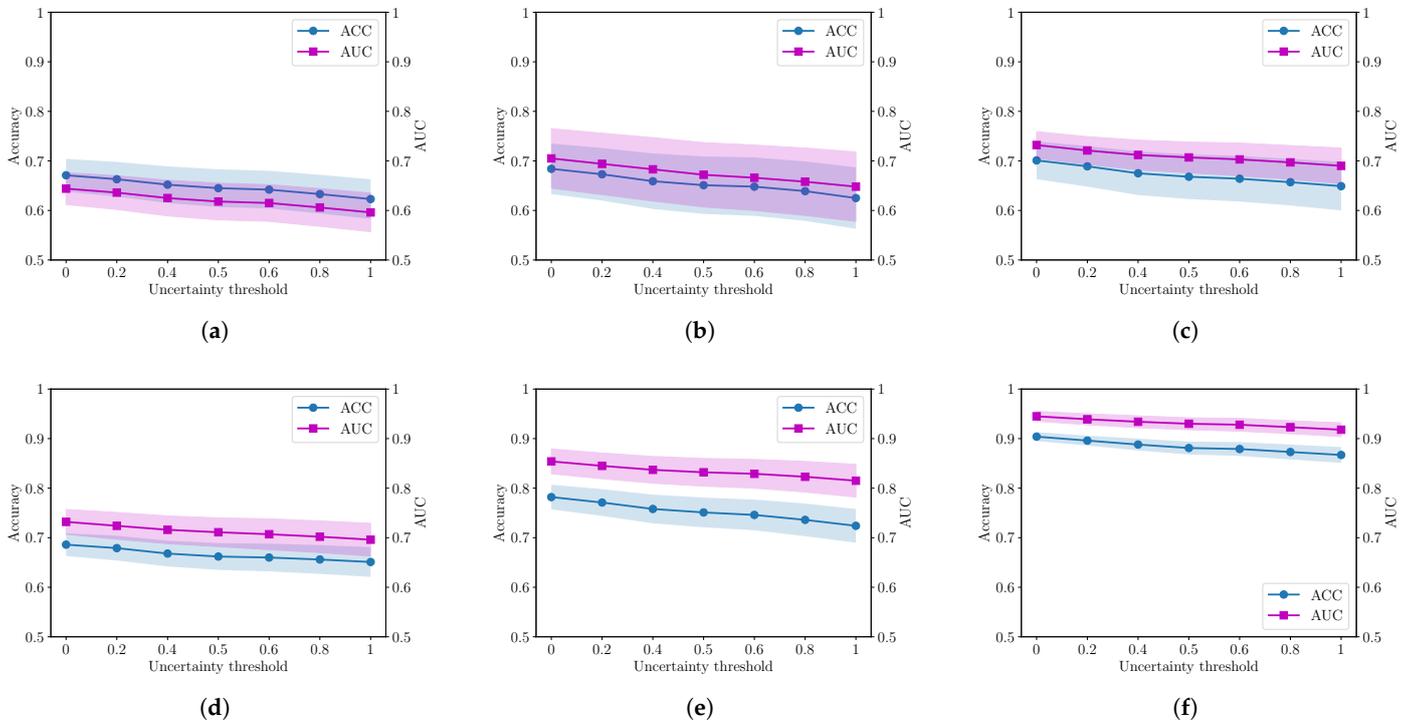
Dataset	Downsampler	SCAB	Metric		
			AUC	ACC	
BreakHis	40×	✓	✓	<b>0.618 ± 0.037</b>	<b>0.645 ± 0.038</b>
		✓		0.594 ± 0.039	0.624 ± 0.040
			✓	0.610 ± 0.038	0.639 ± 0.039
				0.590 ± 0.039	0.622 ± 0.040
	100×	✓	✓	<b>0.672 ± 0.035</b>	<b>0.651 ± 0.058</b>
		✓		0.655 ± 0.046	0.633 ± 0.070
			✓	0.666 ± 0.039	0.646 ± 0.064
				0.652 ± 0.047	0.631 ± 0.072
	200×	✓	✓	<b>0.707 ± 0.031</b>	<b>0.668 ± 0.045</b>
		✓		0.691 ± 0.027	0.643 ± 0.052
			✓	0.700 ± 0.029	0.660 ± 0.049
				0.688 ± 0.027	0.641 ± 0.053
	400×	✓	✓	<b>0.711 ± 0.029</b>	<b>0.662 ± 0.027</b>
		✓		0.693 ± 0.032	0.637 ± 0.029
			✓	0.703 ± 0.030	0.653 ± 0.028
				0.690 ± 0.032	0.635 ± 0.029
Oxford-IIIT Pet	✓	✓	<b>0.832 ± 0.028</b>	<b>0.751 ± 0.030</b>	
	✓		0.821 ± 0.023	0.736 ± 0.029	
		✓	0.828 ± 0.026	0.746 ± 0.029	
			0.818 ± 0.022	0.733 ± 0.029	
Chest X-ray	✓	✓	<b>0.930 ± 0.012</b>	<b>0.881 ± 0.013</b>	
	✓		0.919 ± 0.010	0.871 ± 0.012	
		✓	0.926 ± 0.011	0.878 ± 0.013	
			0.917 ± 0.010	0.869 ± 0.012	

#### 4.4.2. Uncertainty Threshold

The proposed method produces a subjective logic opinion for each psuedo-view, allowing the final prediction to be filtered by an uncertainty threshold. A prediction is accepted only when its fused uncertainty falls below a predefined value. This mechanism enables the model to reject low-confidence predictions and potentially improve reliability. To study the effect of this threshold, the proposed method is evaluated on the BreakHis (40×, 100×, 200×, and 400×), Oxford-IIIT Pet, and Chest X-ray datasets under seven uncertainty thresholds: 0.0, 0.2, 0.4, 0.5, 0.6, 0.8, and 1.0.

Figure 7 illustrates the mean accuracy and AUC, along with their variances, across all datasets. Both metrics exhibit consistent behavior as the threshold varies. When the threshold is set to a very large value (0.8 or 1.0), the model accepts nearly all predictions–

including highly uncertain ones—and therefore yields slightly lower accuracy and AUC. As the threshold decreases, both metrics improve because ambiguous samples are gradually excluded from evaluation. In the mid-range (0.4~0.6), the curves for accuracy and AUC stabilize for all datasets, indicating that the model has filtered out most of the unreliable predictions while still retaining a sufficiently large number of samples.



**Figure 7.** Classification performance on different real-world datasets with varying uncertainty thresholds. (a) BreakHis 40×; (b) BreakHis 100×; (c) BreakHis 200×; (d) BreakHis 400×; (e) Oxford-IIIT Pet; (f) Chest X-ray.

For example, BreakHis 40× shows a rise in AUC from 0.596 at threshold 1.0 to 0.618 at threshold 0.5, accompanied by an accuracy increase from 0.623 to 0.645. Similar improvements are observed in the BreakHis 100×, 200×, and 400× subsets, as well as in Oxford-IIIT Pet and Chest X-ray. The latter dataset, which exhibits lower inherent uncertainty, shows a modest but consistent improvement in both metrics as the threshold decreases.

Although thresholds lower than 0.5, such as 0.2 and 0.0, continue to increase accuracy and AUC, these gains come at the cost of discarding a larger portion of the test predictions. Such aggressive filtering reduces evaluation coverage and biases the reported metrics toward only the easiest samples. Conversely, thresholds higher than 0.5 retain too many unreliable predictions and lead to suboptimal performance.

Considering these observations, the threshold of 0.5 is selected for all experiments in this paper because it represents a balanced point:

- it preserves most test samples;
- it delivers strong and stable performance across all datasets;
- it avoids the coverage loss and evaluation bias associated with overly strict thresholds.

#### 4.4.3. Robustness to Noisy Input

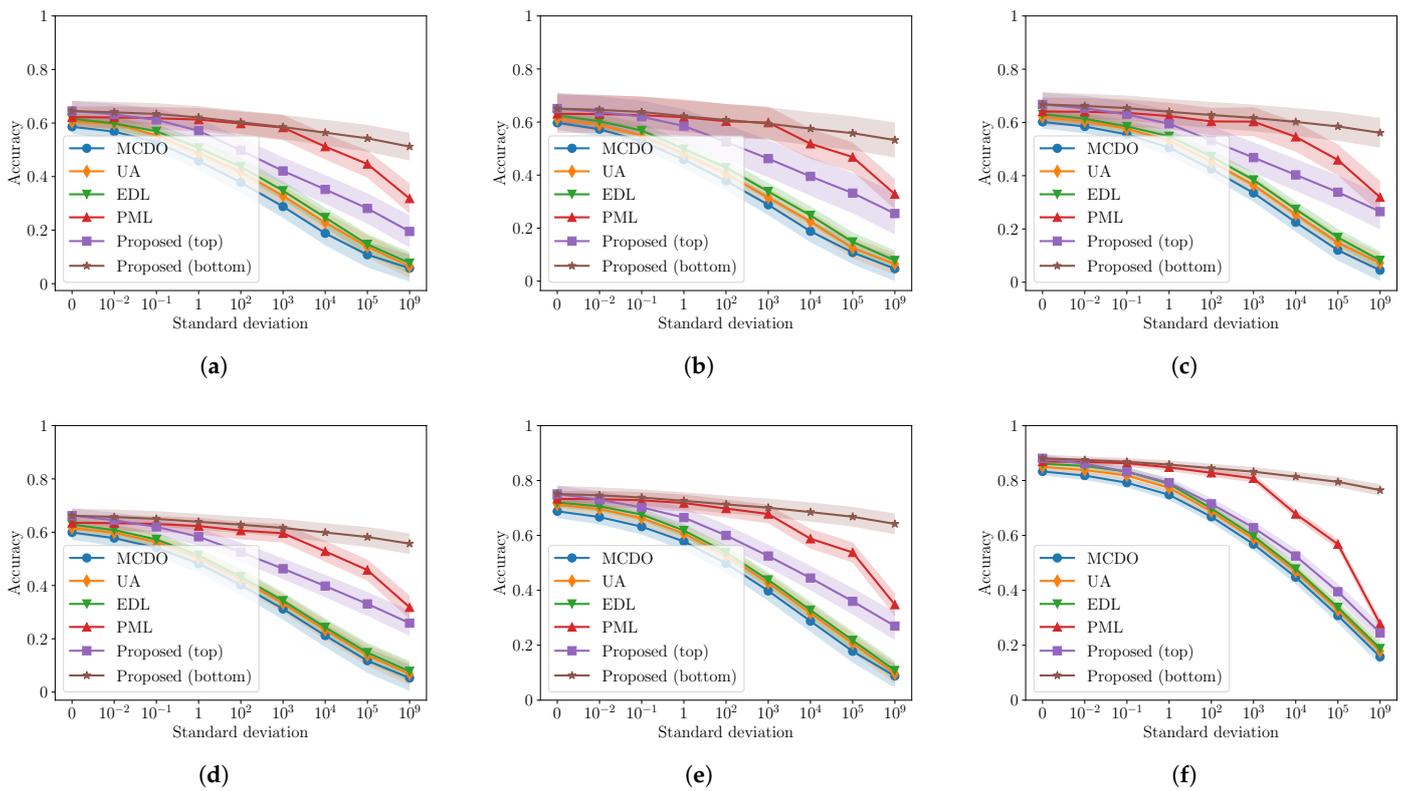
To assess the robustness of the proposed model under degraded input conditions, an ablation study where one of the two psuedo-views is corrupted with Gaussian noise of zero mean and standard deviation ranging from 0 to  $10^9$  is conducted. The experiment is

performed on conventional PML, the proposed multiscale PML (both with two psuedo-views), and three single-view uncertainty-aware baselines: MCDO, UA, and EDL. For the proposed method, two scenarios are tested:

- noise applied to the top branch;
- noise applied to the bottom branch.

In the multiscale grid architecture, the top branch influences all lower branches, whereas the bottom branch cannot affect the upper-level view.

Figure 8 presents the mean accuracy and variance across the BreakHis (40×, 100×, 200×, 400×), Oxford-IIIT Pet, and Chest X-ray datasets. All models perform well when the input is clean ( $\sigma = 0$ ). However, as the noise level increases, the differences among methods become increasingly pronounced.



**Figure 8.** Classification accuracy on different real-world datasets with varying noise levels. (a) BreakHis 40×; (b) BreakHis 100×; (c) BreakHis 200×; (d) BreakHis 400×; (e) Oxford-IIIT Pet; (f) Chest X-ray.

**Comparison with Single-View Methods:** The single-view baselines (MCDO, UA, and EDL) degrade rapidly under noise. MCDO shows the steepest decline, consistent with its known sensitivity to input corruption. For example, on BreakHis 40×, its accuracy drops from 0.586 at  $\sigma = 0$  to 0.458 at  $\sigma = 1$  and further to 0.378 at  $\sigma = 10^2$ . UA exhibits slightly better noise tolerance, while EDL offers the strongest robustness among the three, thanks to its evidence-based modeling. Nevertheless, all single-view methods suffer substantial performance degradation as noise level increases because they rely solely on one noisy feature stream.

**Comparison with Conventional PML:** Conventional PML is more robust than single-view methods at low noise levels due to its multiview design, but its independent branches limit its resilience to extreme noise. When one branch is corrupted, the unaffected branch provides partial compensation, but the absence of cross-scale interaction prevents the

model from redistributing reliable evidence effectively. For example, on BreakHis 200 $\times$ , conventional PML drops from 0.641 at  $\sigma = 0$  to 0.604 at  $\sigma = 10^2$  and 0.546 at  $\sigma = 10^4$ .

**Proposed Method (Bottom-Branch Noise):** When noise is applied to the bottom branch of the proposed method, the upper branch remains intact and continues to dominate the fused opinion. As a result, the model achieves significantly better performance than conventional PML and all single-view baselines. For example, on BreakHis 200 $\times$ , the bottom-noisy curve decreases from 0.668 ( $\sigma = 0$ ) to 0.628 ( $\sigma = 10^2$ ), which is higher than the conventional PML's 0.604, UA's 0.453, and MCDO's 0.425 at the same noise level. This result demonstrates the advantage of the multiscale grid architecture: noise affecting a low-resolution view can be compensated for by clean high-resolution evidence.

**Proposed Method (Top-Branch Noise):** When noise corrupts the top branch, the proposed method exhibits a clear limitation. As the architecture propagates information downward through cross-branch interactions, noise introduced at the highest-resolution branch affects all subsequent views. Consequently, the accuracy deteriorates more sharply than in the bottom-noisy case and may fall below that of conventional PML at higher noise levels. For example, on BreakHis 400 $\times$ , the top-noisy accuracy drops from 0.662 at  $\sigma = 0$  to 0.463 at  $\sigma = 10^3$ , whereas conventional PML remains higher at 0.596. This result confirms that the strong coupling between branches, while beneficial in clean settings, makes the system vulnerable to corruption of the top-level features.

**Summary:** The proposed method delivers superior performance under noisy input conditions when the corruption is applied to the bottom branch, outperforming conventional PML and all single-view uncertainty-aware baselines across datasets and noise levels. However, when the top branch is corrupted, the noise propagates to all lower branches, revealing a structural limitation of the multiscale grid architecture. This trade-off highlights the importance of designing noise-resilient cross-branch interactions and motivates future exploration of selective or gated information flow mechanisms.

## 5. Conclusions

This paper introduced a multiscale grid architecture for pseudo-multiview learning (PML), designed to address the scalability and training stability limitations of conventional PML. By constructing additional psuedo-views through multiscale downsampling and enabling structured cross-branch interactions, the proposed framework extracts complementary features more effectively while controlling parameter growth. Combined with subjective-logic-based opinion fusion, the method offers a principled mechanism for integrating evidence across psuedo-views in an uncertainty-aware manner.

Comprehensive experiments on three datasets from distinct domains—BreakHis for microscopic pathology imaging, Oxford-IIIT Pet for natural-image fine-grained classification, and Chest X-ray for medical radiography—demonstrate that the proposed model achieves consistently higher AUC and accuracy than conventional PML and single-view uncertainty-aware baselines such as Monte Carlo Dropout, uncertainty-aware attention, and evidential deep learning. The model converges more rapidly during training, exhibits more stable optimization behavior, and achieves measurable reductions in parameter count as the number of psuedo-views increases. Ablation studies further show that both the downsampling mechanism and the Spatial-Channel Attention Block (SCAB) contribute to performance improvements, with SCAB providing the most significant gains.

The noise-sensitivity study highlights both the strengths and limitations of the proposed architecture. When noise affects lower-resolution branches, the multiscale design suppresses its impact and preserves performance. However, noise injected into the highest-resolution branch propagates downward through the grid and degrades all psuedo-views, revealing a structural vulnerability associated with top-down feature coupling.

Overall, the proposed multiscale grid architecture provides an effective and efficient refinement of pseudo-multiview learning, offering improved stability, convergence, and performance across diverse datasets. Future work will explore selective or gated cross-branch connections to mitigate noise propagation, develop adaptive view-weighting mechanisms, and extend the framework to additional domains and modalities.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are openly available at BreakHis: <https://iee-dataport.org/documents/breakhis-dataset> (accessed on 7 October 2025); Oxford-IIIT Pet: <https://www.robots.ox.ac.uk/~vgg/data/pets/> (accessed on 7 October 2025); Chest X-ray: <https://data.mendeley.com/datasets/rscbjbr9sj/2> (accessed on 7 October 2025).

**Conflicts of Interest:** The author declares no conflict of interest.

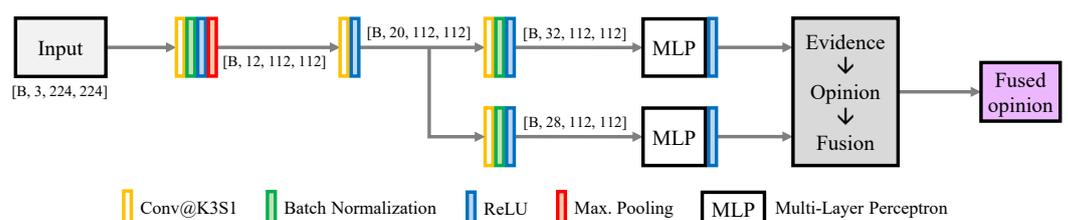
### Appendix A

In pseudo-multiview learning (PML), adding more views increases the dimensionality of the evidence representations and leads to substantial growth in the number of trainable parameters. As a result, training becomes unstable when the number of views becomes too large. This limitation is demonstrated on the Oxford-IIIT Pet dataset [40] by varying the number of psuedo-views from two to six. Assessment metrics include Receiver Operating Characteristic Area Under the Curve (AUC) and Accuracy (ACC). Figure A1 illustrates the architectural design of the model used in this experiment for two views, where  $B$  denotes the batch size. The same architecture is used for three to six views, with additional branches of  $[B, 28, 112, 112]$  added in a similar manner.

Table A1 shows that performance improves from two to four views due to increased feature diversity, but accuracy decreases when using five and six views, indicating that the model fails to converge reliably under the expanded parameter space. Training models with seven or more views results in out-of-memory errors, showing that naive scaling of PML is not feasible.

**Table A1.** Model performance on the Oxford-IIIT Pet dataset for different numbers of psuedo-views.

Number of Views	AUC	ACC
2	0.808 ± 0.029	0.725 ± 0.033
3	0.821 ± 0.031	0.739 ± 0.035
4	0.832 ± 0.028	0.751 ± 0.030
5	0.641 ± 0.037	0.562 ± 0.041
6	0.609 ± 0.045	0.523 ± 0.048
≥7	Out of memory	



**Figure A1.** Architectural design of the model in the pseudo-view-varying experiment, where the  $[B, C, H, W]$  notation denotes the batch size, channel number, height, and width, respectively.

## References

1. Zayed, N.; Eldeeb, G.; Yassine, I. Classification method based on surf and sift features for alzheimer diagnosis using diffusion tensor magnetic resonance imaging. *Sci. Rep.* **2025**, *15*, 9782. [[CrossRef](#)]
2. Long, X.; Hu, S.; Hu, Y.; Gu, Q.; Ishii, I. An FPGA-Based Ultra-High-Speed Object Detection Algorithm with Multi-Frame Information Fusion. *Sensors* **2019**, *19*, 3707. [[CrossRef](#)]
3. Shao, J.; Liu, X.; He, W. Kernel Based Data-Adaptive Support Vector Machines for Multi-Class Classification. *Mathematics* **2021**, *9*, 936. [[CrossRef](#)]
4. Escobar, J.; Rodriguez, F.; Prieto, B.; Kimovski, D.; Ortiz, A.; Damas, M. A distributed and energy-efficient KNN for EEG classification with dynamic money-saving policy in heterogeneous clusters. *Computing* **2023**, *105*, 2487–2510. [[CrossRef](#)]
5. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
6. Tan, M.; Le, Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning (ICML), Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
7. Rao, Y.; Zhao, W.; Tang, Y.; Zhou, J.; Lim, S.; Lu, J. HorNet: Efficient high-order spatial interactions with recursive gated convolutions. In Proceedings of the 36th International Conference on Neural Information Processing Systems (NIPS), New Orleans, LA, USA, 28 November–9 December 2022; pp. 10353–10366.
8. Woo, S.; Debnath, S.; Hu, R.; Chen, X.; Liu, Z.; Kweon, I.S.; Xie, S. ConvNeXt V2: Co-designing and Scaling ConvNets with Masked Autoencoders. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 16133–16142. [[CrossRef](#)]
9. Khan, A.; Khan, A. Multi-axis vision transformer for medical image segmentation. *Eng. Appl. Artif. Intell.* **2025**, *158*, 111251. [[CrossRef](#)]
10. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Very Deep Convolutional Networks for Large-Scale Image Recognition. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [[CrossRef](#)]
11. Cengil, E.; Cinar, A. The effect of deep feature concatenation in the classification problem: An approach on COVID-19 disease detection. *Int. J. Imaging Syst. Technol.* **2021**, *32*, 26–40. [[CrossRef](#)]
12. Guo, C.; Pleiss, G.; Sun, Y.; Weinberger, K. On calibration of modern neural networks. In Proceedings of the 34th International Conference on Machine Learning (ICML), Sydney, Australia, 6–11 August 2017; pp. 1321–1330.
13. Kwon, Y.; Won, J.; Kim, B.; Paik, M. Uncertainty quantification using Bayesian neural networks in classification: Application to biomedical image segmentation. *Comput. Stat. Data Anal.* **2020**, *142*, 106816. [[CrossRef](#)]
14. Linka, K.; Holzapfel, G.; Kuhl, E. Discovering uncertainty: Bayesian constitutive artificial neural networks. *Comput. Methods Appl. Mech. Eng.* **2025**, *433*, 117517. [[CrossRef](#)]
15. Amersfoort, J.; Smith, L.; Teh, Y.; Gal, Y. Uncertainty estimation using a single deep deterministic neural network. In Proceedings of the 37th International Conference on Machine Learning (ICML), Vienna, Austria, 12–18 July 2020; pp. 9690–9700.
16. Alrweili, H.; Alotaibi, E. Bayesian and non-bayesian estimation of Marshall-Olkin XLindley distribution in presence of censoring, cure fraction, and application on medical data. *Alex. Eng. J.* **2025**, *112*, 633–646. [[CrossRef](#)]
17. Jøsang, A. *Subjective Logic: A Formalism for Reasoning Under Uncertainty*; Springer: Cham, Switzerland, 2016.
18. Ngo, D. Pseudo-Multiview Learning Using Subjective Logic for Enhanced Classification Accuracy. *Mathematics* **2025**, *13*, 2085. [[CrossRef](#)]
19. Li, Y.; Yang, M.; Zhang, Z. A Survey of Multi-View Representation Learning. *IEEE Trans. Knowl. Data Eng.* **2019**, *31*, 1863–1883. [[CrossRef](#)]
20. Zhang, C.; Adeli, E.; Zhou, T.; Shen, X.; Shen, D. Multi-Layer Multi-View Classification for Alzheimer’s Disease Diagnosis. In Proceedings of the 32nd AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; pp. 4406–4413.
21. Zhang, C.; Cui, Y.; Han, Z.; Zhou, J.T.; Fu, H.; Hu, Q. Deep Partial Multi-View Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 2402–2415. [[CrossRef](#)]
22. Long, T.; Xie, Y.; Chen, X.; Zhang, W.; Cao, Q.; Yu, Y. Multi-View Graph Representation for Programming Language Processing: An Investigation into Algorithm Detection. In Proceedings of the 36th AAAI Conference on Artificial Intelligence, Online, 22 February–1 March 2022; pp. 5792–5799. [[CrossRef](#)]
23. Liu, X.; Zhu, X.; Li, M.; Wang, L.; Tang, C.; Yin, J.; Shen, D.; Wang, H.; Gao, W. Late Fusion Incomplete Multi-View Clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 2410–2423. [[CrossRef](#)]
24. Liu, X.; Li, M.; Tang, C.; Xia, J.; Xiong, J.; Liu, L.; Kloft, M.; Zhu, E. Efficient and Effective Regularized Incomplete Multi-View Clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 2634–2646. [[CrossRef](#)]
25. Liu, X. Hyperparameter-Free Localized Simple Multiple Kernel K-means With Global Optimum. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 8566–8576. [[CrossRef](#)] [[PubMed](#)]
26. Wang, J.; Tang, C.; Zheng, X.; Liu, X.; Zhang, W.; Zhu, E.; Zhu, X. Fast Approximated Multiple Kernel K-Means. *IEEE Trans. Knowl. Data Eng.* **2024**, *36*, 6171–6180. [[CrossRef](#)]

27. Tang, C.; Li, Z.; Wang, J.; Liu, X.; Zhang, W.; Zhu, E. Unified One-Step Multi-View Spectral Clustering. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 6449–6460. [[CrossRef](#)]
28. Hermosilla, D.; Codorniu, R.; Baracaldo, R.; Zamora, R.; Rodriguez, D.; Mayor, J.; Alvarez, J. Monte Carlo Dropout for Uncertainty Estimation and Motor Imagery Classification. *Sensors* **2021**, *21*, 7241. [[CrossRef](#)]
29. Heo, J.; Lee, H.; Kim, S.; Lee, J.; Kim, K.; Yang, E.; Ju, S. Uncertainty-aware attention for reliable interpretation and prediction. In Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS), Montreal, QC, Canada, 3–8 December 2018; pp. 917–926.
30. Sensoy, M.; Kaplan, L.; Kandemir, M. Evidential deep learning to quantify classification uncertainty. In Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS), Montreal, QC, Canada, 3–8 December 2018; pp. 3183–3193.
31. Chen, T.; Komblith, S.; Norouzi, M.; Hinton, G. A simple framework for contrastive learning of visual representations. In Proceedings of the 37th International Conference on Machine Learning (ICML), Virtual, 13–18 July 2020; pp. 1597–1607.
32. Liu, M.; Breuel, T.; Kautz, J. Unsupervised image-to-image translation networks. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017; pp. 700–708.
33. Gidaris, S.; Singh, P.; Komodakis, N. Unsupervised Representation Learning by Predicting Image Rotations. In Proceedings of the 6th International Conference on Learning Representations (ICLR), Vancouver, BC, Canada, 30 April–3 May 2018.
34. Huang, Y.; Shen, Z.; Li, T.; Lv, F. Unified view imputation and feature selection learning for incomplete multi-view data. In Proceedings of the 33rd International Joint Conference on Artificial Intelligence (IJCAI), Jeju, Republic of Korea, 3–9 August 2024; pp. 4192–4200. [[CrossRef](#)]
35. Zamir, S.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.; Yang, M.; Shao, L. Learning Enriched Features for Real Image Restoration and Enhancement. In Proceedings of the 16th European Conference on Computer Vision (ECCV), Glasgow, UK, 23–28 August 2020; pp. 492–511. [[CrossRef](#)]
36. Wu, H.; Li, Z.; Wen, Z. MGFNet: A multiscale gated fusion network for multimodal semantic segmentation. *Vis. Comput.* **2025**, *41*, 9043–9055. [[CrossRef](#)]
37. Huo, X.; Sun, G.; Tian, S.; Wang, Y.; Yu, L.; Long, J.; Zhang, W.; Li, A. HiFuse: Hierarchical multi-scale feature fusion network for medical image classification. *Biomed. Signal Process. Control* **2024**, *87*, 105534. [[CrossRef](#)]
38. Liu, X.; Shi, Z.; Wu, Z.; Chen, J.; Zhai, G. GridDehazeNet+: An Enhanced Multi-Scale Network With Intra-Task Knowledge Transfer for Single Image Dehazing. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 870–884. [[CrossRef](#)]
39. Spanhol, F.A.; Oliveira, L.S.; Petitjean, C.; Heutte, L. A Dataset for Breast Cancer Histopathological Image Classification. *IEEE Trans. Biomed. Eng.* **2015**, *63*, 1455–1462. [[CrossRef](#)] [[PubMed](#)]
40. Parkhi, O.M.; Vedaldi, A.; Zisserman, A.; Jawahar, C.V. Cats and dogs. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 16–21 June 2012; pp. 3498–3505. [[CrossRef](#)]
41. Kermany, D.S.; Goldbaum, M.; Cai, W.; Valentim, C.C.; Liang, H.; Baxter, S.L.; McKeown, A.; Yang, G.; Wu, X.; Yan, F.; et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell* **2018**, *172*, 1122–1131. [[CrossRef](#)]
42. Zhou, K.; Liu, Z.; Qiao, Y.; Xiang, T.; Loy, C.C. Domain Generalization: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 4396–4415. [[CrossRef](#)]
43. Wang, J.; Lan, C.; Liu, C.; Ouyang, Y.; Qin, T.; Lu, W.; Chen, Y.; Zeng, W.; Yu, P. Generalizing to Unseen Domains: A Survey on Domain Generalization. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 8052–8072. [[CrossRef](#)]
44. Microsoft. Neural Network Intelligence. Available online: <https://nni.readthedocs.io/en/stable/> (accessed on 17 December 2024).

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