

## Article

# Weighted Sampling with Frequency-Aware Spatial Attention for Imbalanced Image Classification

Shiqi Zhang <sup>1,\*</sup> and Peng Li <sup>2</sup> <sup>1</sup> School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081, China<sup>2</sup> Department of Earth and Planetary Sciences, Stanford University, Stanford, CA 94305, USA; pli6@stanford.edu

\* Correspondence: 1120231299@bit.edu.cn; Tel.: +86-15202361744

## Abstract

Class imbalance remains a critical challenge in image classification, where underrepresented classes often receive insufficient training attention and exhibit poor recognition performance. In this study, we propose a hybrid framework that combines weighted sampling with frequency-aware spatial attention (WSFSA) to address class imbalance at both the data and feature levels. The weighted sampler improves the training exposure of minority classes, while the frequency-aware spatial attention module incorporates class-frequency information into spatial attention to enhance discriminative feature responses for underrepresented classes. We evaluate the proposed method on four MedMNIST benchmarks, DermaMNIST, BloodMNIST, OrganCMNIST, and DermaMNIST-224, using a ResNet-18 backbone. Results show that WFSFA provides the clearest benefit on the severely imbalanced DermaMNIST and DermaMNIST-224 datasets, performing comparably to the strongest baseline methods while showing particular benefits under severe class imbalance. On OrganCMNIST, WFSFA provides moderate gains, while on BloodMNIST, where the imbalance effect is weaker, all methods perform similarly. Per-class analysis further shows that WFSFA improves sensitivity for several minority or difficult classes while maintaining high specificity across most classes. These findings suggest that combining sampling-level and feature-level rebalancing is a practical strategy for improving class-balanced recognition, particularly under severe class imbalance.

**Keywords:** frequency attention; weighted sampler; hybrid neural network; imbalanced classification

Academic Editors: Jyh-Cheng Chen  
and Kuangyu Shi

Received: 11 May 2026

Revised: 24 June 2026

Accepted: 30 June 2026

Published: 1 July 2026

**Copyright:** © 2026 by the authors.  
Licensee MDPI, Basel, Switzerland.  
This article is an open access article  
distributed under the terms and  
conditions of the [Creative Commons  
Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

## 1. Introduction

### 1.1. Overall Deep Learning Methods for Image Classification

The task of image classification serves as a fundamental application of deep neural networks (DNNs), with broad use across numerous downstream tasks such as object segmentation [1–3], anomaly detection [4–6], and multi-class classification [7–9]. The technique of image classification has been widely adopted across diverse applications, including autonomous driving, medical imaging, remote sensing, manufacturing, agriculture, etc. Despite the remarkable progress of deep learning in computer vision and imaging, class imbalance remains a critical challenge. In real-world applications, class imbalance arises naturally from the inherent rarity or uneven occurrence of certain conditions or phenomena. However, these underrepresented classes can be critical for understanding the underlying problem and may carry important practical significance. For example, in medical imaging,

datasets often contain thousands of normal scans but only a limited number of samples are depicted as rare cancers, uncommon lesions, or early-stage diseases [10,11], and these limited samples are our target of interests that we want to study. Similar patterns appear across multiple domains, such as industrial defect detection [12,13], environmental monitoring [14], and even disaster prediction [15,16], where abnormal events are by nature infrequent but very important to identify accurately.

A wide range of deep learning approaches have been proposed to address the problem of imbalance classification. Here we broadly group the methods into data-level, algorithm-level, and feature-level strategies. Data-level methods, such as oversampling [17,18], undersampling [19,20], and class-balanced sampling [21], directly modify the training distribution to equalize class occurrence. These techniques are straightforward, but excessive oversampling can lead to overfitting, while undersampling may discard valuable information from majority classes. Meanwhile, algorithm-level strategies tackle imbalance by modifying the learning objective. The example methods include class-weighted cross-entropy [22], focal loss [23,24], and cost-sensitive optimization [25,26]. Such methods can emphasize minority classes during training, but they often introduce additional hyper-parameters and can destabilize optimization. Moreover, feature-level approaches aim to enhance the model's representational power for rare classes, often through attention mechanisms [27–29], feature reweighting [30], or contrastive learning [31–33]. These methods adaptively focus on discriminative features but typically rely on spatial or channel saliency without explicitly considering the statistical frequency of each class. Here we will introduce in detail the methods in each category.

### *1.2. Data-Level Strategies for Imbalanced Learning*

Most of the data-level methods are developed to mitigate class imbalance by operating either on the data or the learning objective. Data-level strategies aim to rebalance the class distribution in the training set. Common techniques include oversampling of minority-class samples [17,18], undersampling of majority-class samples [19], and synthetic data generation (e.g., SMOTE or GAN-based augmentation) [34,35]. These methods ensure that minority classes are sufficiently represented during training, which can improve the model's sensitivity to rare events. However, oversampling can lead to overfitting due to repeated exposure to the same minority samples, while undersampling may discard valuable information from majority classes. Weighted sampling schemes, which assign probabilities inversely proportional to class frequency, provide a compromise by emphasizing minority classes without fully discarding majority data.

### *1.3. Algorithm-Level Strategies for Imbalanced Learning*

Algorithm-level strategies modify the loss function or optimization objective to emphasize the contribution of minority classes. Examples include class-weighted cross-entropy [22], focal loss [23,24] and cost-sensitive learning [25,26], all of which amplify the gradient contributions of underrepresented classes. While effective in some scenarios, these approaches can introduce additional hyper-parameters and sometimes destabilize training, particularly when combined with data-level methods. In practice, careful tuning is required to balance minority-class emphasis with overall model stability.

Although both data-level and algorithm-level approaches have shown promise, they often operate independently of the model's feature representation. As a result, minority-class features may still be underrepresented in the learned latent space, especially in complex medical imaging tasks with subtle or localized pathologies. This limitation motivates hybrid approaches that integrate feature-level adaptation, such as attention mechanisms, with data-level rebalancing.

#### 1.4. Attention Mechanisms for Imbalanced Classification

Attention mechanisms have become a fundamental component in modern deep learning architectures, enabling models to focus selectively on informative regions or features while suppressing irrelevant background information. In the context of imbalanced classification, attention provides an effective way to amplify the representation of minority classes without explicitly altering the data distribution. Early works primarily adopted spatial and channel-wise attention to enhance feature discrimination [36,37]. Spatial attention highlights localized regions that are critical for classification decisions, which are especially beneficial in medical imaging where pathological features may occupy small or subtle regions [38,39]. Channel attention, on the other hand, assigns adaptive weights to feature channels, allowing the network to prioritize class-relevant feature maps [40,41]. Popular frameworks such as the Squeeze-and-Excitation (SE) block [42], CBAM (Convolutional Block Attention Module) [43], and self-attention mechanisms in Vision Transformers (ViTs) [44] have demonstrated improved robustness and interpretability under moderate-imbalance conditions.

However, most existing attention modules are based on feature reweighting. They enhance features purely based on activation strength or spatial saliency, without considering how frequently each class appears in the dataset. As a result, these mechanisms may still favor majority-class patterns, especially when the model learns biased representations during training. Some recent studies have attempted to introduce class-aware attention [45] or reweighting schemes [46], but these approaches often rely on external supervision or predefined priors rather than leveraging class-frequency information directly.

#### 1.5. Hybrid and Frequency-Aware Approaches

Recent studies suggest that data-level or algorithm-level strategies alone may not fully address class imbalance, especially in medical imaging, where minority-class features can be subtle and localized. Hybrid approaches that combine multiple levels of rebalancing therefore have been proposed [47,48]. Such methods typically integrate data-level sampling with feature-level adaptation, enabling the model to both see rare classes more frequently and allocate representational capacity to their distinctive features. Several recent studies have explored class-aware or cost-sensitive attention mechanisms, demonstrating improved performance on long-tailed or imbalanced datasets. However, many of these approaches either require external supervision to define class importance or rely on static heuristics that do not adapt dynamically during training. Furthermore, most prior work focuses on either feature-level adaptation or sampling strategies individually, leaving the potential synergy between the two largely unexplored.

#### 1.6. Scope of the Work

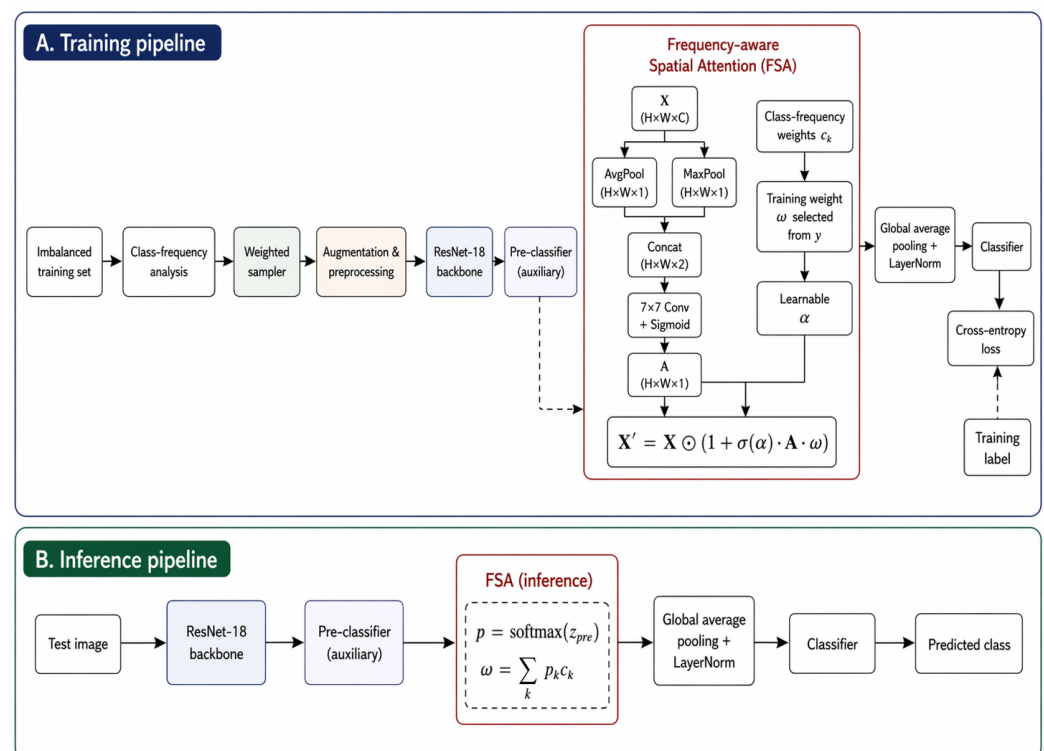
Motivated by the above observations, our work develops a unified framework that combines a weighted sampling with frequency-aware spatial attention scheme (WSFSA). Our frequency-aware spatial attention module (FSA) embeds class-frequency information into the attention process, allowing the model to focus adaptively on features from minority classes. Combined with a weighted sampler (WS) that balances training data exposure, this framework unifies rebalancing and frequency-guided feature learning to overcome key limitations of existing imbalance mitigation techniques. This design simultaneously addresses sample-level exposure and feature-level emphasis, providing a complementary and interpretable solution to the challenges of imbalanced image classification.

The structure of this paper is organized as follows. Section 1 presents the introduction and background work on class imbalance mitigation using deep learning method. Section 2 introduces the mathematical foundation of proposed framework, including the frequency-

aware spatial attention module and the weighted sampling strategy, as well as the experimental setup, datasets, and evaluation protocols used to assess the proposed method. Section 3 reports and discusses the experimental results, highlighting the effectiveness of the proposed approach compared with existing techniques. Finally, Section 4 concludes the paper with a summary of key findings and potential directions for future research.

## 2. Materials and Methods

In this section, we present the proposed framework for class-imbalanced image classification, which combines frequency-aware spatial attention with a weighted sampling scheme to address both feature- and data-level imbalance. The overall pipeline is illustrated in Figure 1.



**Figure 1.** Overall workflow of the proposed method.

### 2.1. Overview

Our framework is built on a ResNet-18 backbone pretrained on ImageNet. To mitigate class imbalance, we integrate two complementary components that operate at different stages of the learning pipeline. The first component is WS (data-level rebalancing), where a square-root inverse-frequency sampler selects training samples with probability proportional to  $1/\sqrt{f_c}$ . This increases minority-class exposure during training without discarding majority-class data. The second component is FSA (feature-level rebalancing). FSA generates a spatial attention map from channel-wise average pooling and max pooling, and then modulates the backbone feature maps using both local spatial activation and class-frequency weights, defined as  $(1/(f_c + \epsilon))^\beta$ . During training, frequency weights are assigned using ground-truth labels, whereas during inference, an expected frequency weight is estimated from the softmax output of a lightweight pre-classifier. A learnable residual gating parameter  $\alpha$  controls the modulation strength and supports stable gradient flow. Together, weighted sampling and FSA form a hybrid rebalancing mechanism that

addresses imbalance at both the sampling-distribution and feature-representation levels, improving minority-class sensitivity while preserving overall discriminative performance.

## 2.2. Weighted Sampling

Let the training dataset be

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N, \quad y_i \in \{1, \dots, C\} \quad (1)$$

where  $N$  is the total number of samples and  $C$  is the number of classes. Let  $n_c$  denote the number of samples belonging to class  $c$ . To mitigate class imbalance, we assign a sampling weight to each class that is inversely proportional to the square root of its frequency. Specifically, the weight for class  $c$  is defined as

$$w_c = \frac{1}{\sqrt{n_c + \epsilon}} \quad (2)$$

where  $\epsilon > 0$  denotes a small positive constant introduced to avoid division by zero. Each sample  $i$  inherits the weight of its corresponding class:

$$w_i = w_{y_i} \quad (3)$$

The sampling probability of sample  $i$  is obtained by normalizing its weight over all samples:

$$\pi_i = \frac{w_i}{\sum_{j=1}^N w_j} \quad (4)$$

where  $\pi_i$  denotes the probability of sampling sample  $i$ . The class-level sampling probability  $P(c)$  is the aggregate probability that a randomly drawn sample belongs to class  $c$ :

$$P(c) = \sum_{i: y_i=c} \pi_i \quad (5)$$

Substituting  $\pi_i$  from above, the expected sampling probability for class  $c$  becomes

$$P(c) = \frac{n_c w_c}{\sum_{k=1}^C n_k w_k} \quad (6)$$

This formulation increases the likelihood of selecting samples from rare classes, while preventing overcompensation compared to fully inverse-frequency sampling. These weights are normalized, ensuring that minority-class samples are more likely to be selected during training, while reducing the risk of overfitting associated with naive oversampling.

## 2.3. Frequency-Aware Spatial Attention

The FSA module enhances feature-level representations for underrepresented classes. Given a feature map  $X \in \mathbb{R}^{B \times C \times H \times W}$ , we first compute spatial attention using both average pooling and max pooling along the channel dimension:

$$A_{\text{spatial}} = \sigma(\text{Conv2D}([\text{AvgPool}(X), \text{MaxPool}(X)])) \quad (7)$$

where  $\sigma(\cdot)$  denotes the sigmoid activation function. To incorporate class-frequency information, we define a frequency weight for each class:

$$a_c = \frac{1/f_c^\beta}{\max_k (1/f_k^\beta)} \quad (8)$$

where  $f_c$  denotes the empirical frequency of class  $c$ , and  $\beta > 0$  controls the strength of the frequency correction. The attention-modulated feature map is then computed as

$$X' = X \odot \left(1 + \alpha \cdot A_{\text{spatial}} \cdot a_c\right) \quad (9)$$

where  $\alpha$  is a learnable parameter that controls the overall contribution of the attention mechanism, and  $\odot$  denotes element-wise multiplication. This formulation ensures that minority-class features receive proportionally higher attention, thereby improving their discriminative representation.

#### 2.4. Weighted Sampling with Frequency-Aware Spatial Attention

The FSA module is inserted after the final convolutional layer of the ResNet backbone. During training, the WS module selects batches with class-balanced exposure, while the FSA module adaptively emphasizes spatial regions relevant to underrepresented classes. The resulting feature maps are then flattened and passed through fully connected layers for classification. Optionally, a logit-level bias can be added to further correct for class imbalance.

The expected contribution of class  $c$  to the model per training step can be written as

$$\mathbb{E}[Y_c] \propto \pi_c a_c M_c = \frac{f_c^{-\beta_s}}{\sum_{k=1}^C f_k^{-\beta_s}} \cdot \frac{f_c^{-\beta_a}}{\max_k f_k^{-\beta_a}} \cdot M_c \quad (10)$$

where  $f_c$  is the empirical frequency of class  $c$ ,  $\beta_s, \beta_a > 0$  control the strength of sampling and attention reweighting, respectively,  $\pi_c$  is the expected sampling probability of class  $c$ ,  $W_c$  is the frequency weight for attention, and  $M_c$  is the spatial attention map for class  $c$ . This formulation shows how sampling frequency and feature attention are jointly used to prioritize rare classes.

By jointly leveraging data-level sampling and frequency-aware feature modulation, the proposed framework provides a complementary and interpretable approach to imbalanced image classification. The WS module ensures sufficient exposure to minority classes, while FSA amplifies their feature representation in the network, addressing limitations of prior attention-only or sampling-only strategies.

#### 2.5. Experiments Design

In this section, we evaluate the proposed WSFSA method on class-imbalanced image classification.

##### 2.5.1. Dataset

We evaluate the proposed method on four datasets from the MedMNIST v2 collection [49], covering three imaging modalities and different levels of class imbalance. The datasets were deliberately selected to span a range of imbalance factors, from approximately 10× to 58×, enabling evaluation under mild, moderate, and severe class imbalance settings. All datasets use the official predefined training, validation, and test splits to ensure reproducibility. For the three standard-resolution datasets, the original images are  $28 \times 28$  pixels and are resized to  $96 \times 96$  pixels to provide sufficient spatial resolution for convolutional feature extraction. In addition, we include DermaMNIST-224, a higher-resolution version of our most imbalanced benchmark, to further evaluate the generalizability of the proposed method under a more realistic image-resolution setting.

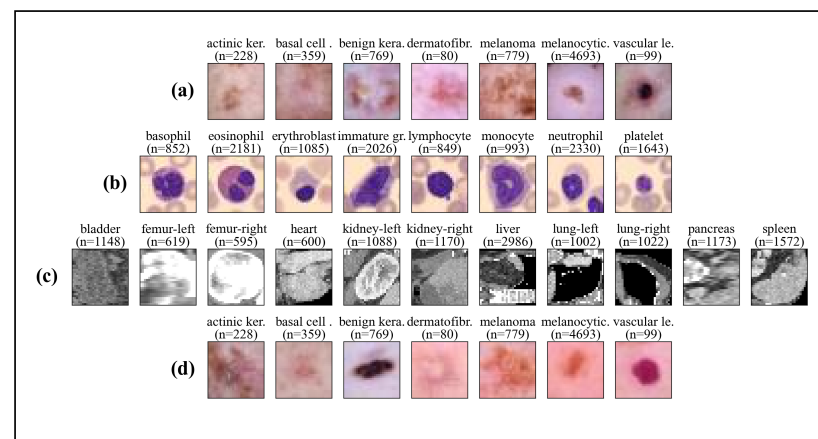
DermaMNIST is a dermatoscopic image classification benchmark comprising 10,015 images across seven diagnostic categories (Figure 2a): actinic keratoses ( $n = 228$ ), basal cell carcinoma ( $n = 359$ ), benign keratosis ( $n = 769$ ), dermatofibroma ( $n = 80$ ),

melanoma ( $n = 779$ ), melanocytic nevi ( $n = 4693$ ), and vascular lesions ( $n = 99$ ). This dataset exhibits severe class imbalance, with an imbalance factor of approximately  $58\times$  (Figure 3a). The train/validation/test split is 7007/1003/2005.

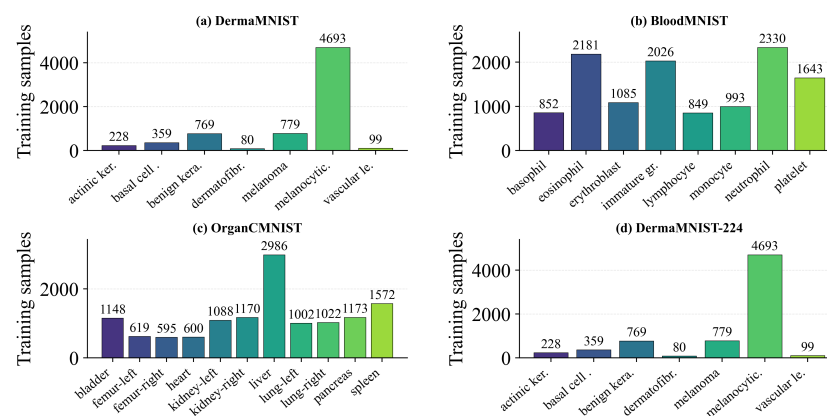
BloodMNIST comprises 17,092 microscopy images of blood cells across eight classes (Figure 2b): basophil, eosinophil, erythroblast, immature granulocyte, lymphocyte, monocyte, neutrophil, and platelet. The imbalance factor is approximately  $10\times$  (Figure 3b), and the train/validation/test split is 11,959/1712/3421. This dataset represents the mildest imbalance among the three, allowing us to assess whether the proposed method remains effective under moderate class imbalance.

OrganCMNIST contains 23,660 coronal-view abdominal CT images from 11 organ classes (Figure 2c): bladder, femur-left, femur-right, heart, kidney-left, kidney-right, liver, lung-left, lung-right, pancreas, and spleen. The imbalance factor is approximately  $13\times$  (Figure 3c), and the train/validation/test split is 13,000/2392/8268. This dataset introduces additional classification difficulty because several anatomically symmetric organ pairs, such as the left and right femur or kidney, have similar visual appearances at low resolution.

DermaMNIST-224 is the higher-resolution version of DermaMNIST (Figure 2d), containing the same images, diagnostic categories, class distribution (Figure 3d), and train/validation/test split. The main difference is that images are provided at  $224 \times 224$  pixels, allowing us to evaluate the proposed method under a higher-resolution setting while retaining the same severe class imbalance of approximately  $58\times$ .



**Figure 2.** Illustration of images and labels of four datasets: (a) DermaMNIST, (b) BloodMNIST, (c) OrganCMNIST, and (d) DermaMNIST-224.



**Figure 3.** Histograms of data samples from four datasets: (a) DermaMNIST, (b) BloodMNIST, (c) OrganCMNIST, and (d) DermaMNIST-224.

Together, these datasets cover dermatoscopy, computed tomography, and microscopy, with imbalance factors ranging from  $10\times$  to  $58\times$  and class counts ranging from 7 to 11. They provide a diverse evaluation setting for assessing the robustness and generalizability of the proposed method across different medical imaging conditions.

### 2.5.2. Network Backbone

We adopt ResNet-18 [50] pretrained on ImageNet as the feature extraction backbone. The final average pooling and fully connected layers are removed, yielding a convolutional feature extractor that produces spatial feature maps  $F$  of dimension  $512 \times H \times W$ , where  $H$  and  $W$  denote the spatial dimensions after downsampling. For the baseline model (Vanilla), a global average pooling (GAP) layer reduces the feature maps to a 512-dimensional vector, which is then projected to the number of classes via a single linear layer.

### 2.5.3. Compared Methods

We compare eight configurations to isolate the contribution of each component:

- Vanilla: ResNet-18 baseline trained with cross-entropy loss and uniform random sampling. This serves as the standard reference.
- Focal: Same architecture as Vanilla, but trained with focal loss,  $(1 - p_t)^\gamma \cdot \text{CE}$ , with  $\gamma = 2.0$ . This setting downweights well-classified examples and focuses learning on hard samples. This serves as the benchmark technology used in literature.
- WS: Vanilla architecture trained with a square-root inverse-frequency weighted random sampler, where each sample's drawing probability is proportional to  $1/\sqrt{f_c}$ . This rebalances the effective training distribution without modifying the loss function.
- CBAM: ResNet-18 augmented with the Convolutional Block Attention Module (CBAM), trained with uniform sampling and cross-entropy loss. This serves as an attention-based baseline for comparison with the proposed FSA module.
- LDAM-DRW: ResNet-18 trained with the Label-Distribution-Aware Margin loss and deferred reweighting strategy. This serves as a representative loss-level method for long-tailed and imbalanced classification.
- LogitAdjust: Vanilla architecture trained with cross-entropy loss and uniform random sampling, followed by post hoc logit adjustment using class-prior information. This evaluates whether class-prior correction at the output level can improve imbalanced classification.
- FSA: ResNet-18 augmented with the proposed frequency-aware spatial attention module, trained with uniform sampling and cross-entropy loss. Logit bias correction is enabled.
- WSFSA: The combined approach using both WS and FSA. The logit bias is disabled to prevent double correction, since the WS module already adjusts the training distribution.

### 2.5.4. Training Details

All models are trained for 30 epochs with a batch size of 128 using the AdamW optimizer with an initial learning rate of  $10^{-4}$  and a weight decay of 0.05. The learning rate follows a cosine annealing schedule with linear warmup over the first 5 epochs. Cross-entropy loss with label smoothing of 0.1 is used for all methods except Focal, which employs focal loss [23] with  $\gamma = 2.0$ .

### 2.5.5. Evaluation Metrics

In our case, we use the following metrics to provide a comprehensive assessment of the proposed models and their performance under class imbalance:

- Overall accuracy: It defines the fraction of correctly classified samples across all classes. This metric is commonly reported, but it can be dominated by majority classes under imbalanced data distributions.
- Macro accuracy (balanced accuracy): It defines the unweighted mean of per-class recall, giving equal importance to each class regardless of its prevalence.
- Macro F1-score: This represents the unweighted mean of per-class F1-scores, balancing precision and recall across all classes.
- AUC-ROC (macro, one-vs-rest): The area under the receiver operating characteristic curve, averaged across all classes using the one-vs-rest strategy and computed from softmax probabilities.
- Per-class sensitivity (recall): It defines the true-positive rate for each class, measuring the fraction of actual positive samples that are correctly identified. This is our key metric because it reflects the model's ability to detect each condition.
- Per-class specificity: It represents the true-negative rate for each class, measuring the fraction of actual negative samples that are correctly rejected.

In our case, we emphasize that macro accuracy, macro F1-score, and per-class sensitivity as the primary metrics of interest, since these metrics better reflect performance on minority classes that may be masked by overall accuracy.

### 3. Results

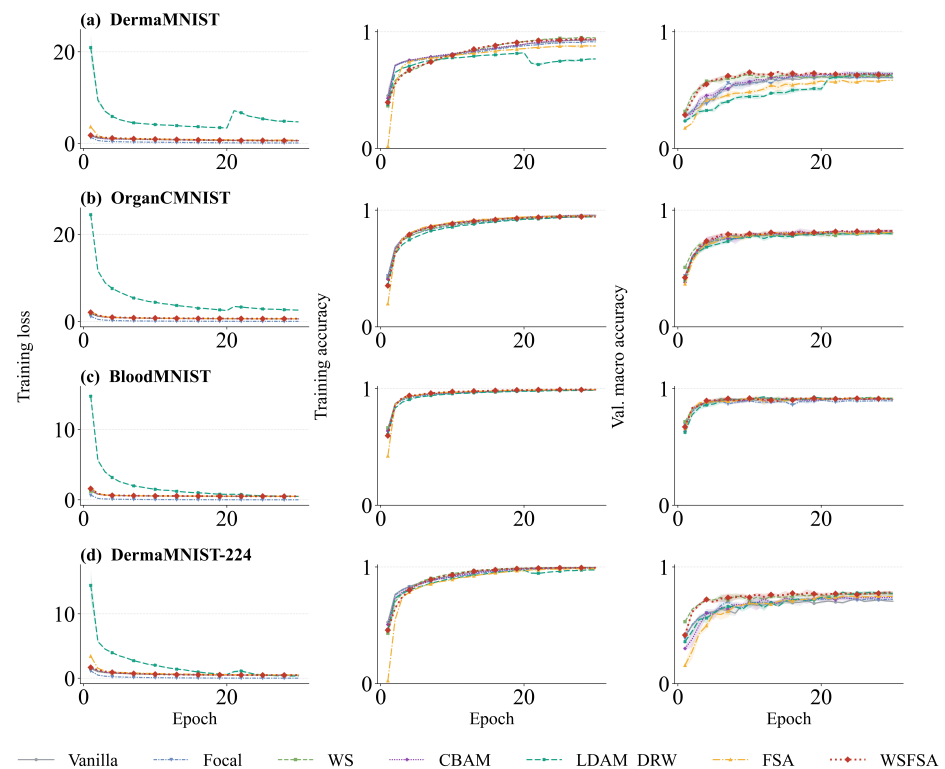
#### 3.1. Training Performance

We conducted all experiments using the training settings described in Section 2.5.4. To improve statistical reliability, each configuration was independently run three times using different random seeds, and the reported results are summarized using the mean and standard deviation. Figure 4 shows the average training loss, training accuracy, and validation macro accuracy over 30 epochs for DermaMNIST, BloodMNIST, OrganCMNIST, and DermaMNIST-224. Across all datasets, most methods converge within the first 20 epochs, with training loss decreasing rapidly in the early stage and then stabilizing. Focal loss reaches the lowest absolute training loss, although its value is not directly comparable with cross-entropy-based losses due to the focal modulation term. LogitAdjust is applied post hoc to a trained baseline rather than trained as a separate model, so it has no training curve and is omitted from this figure.

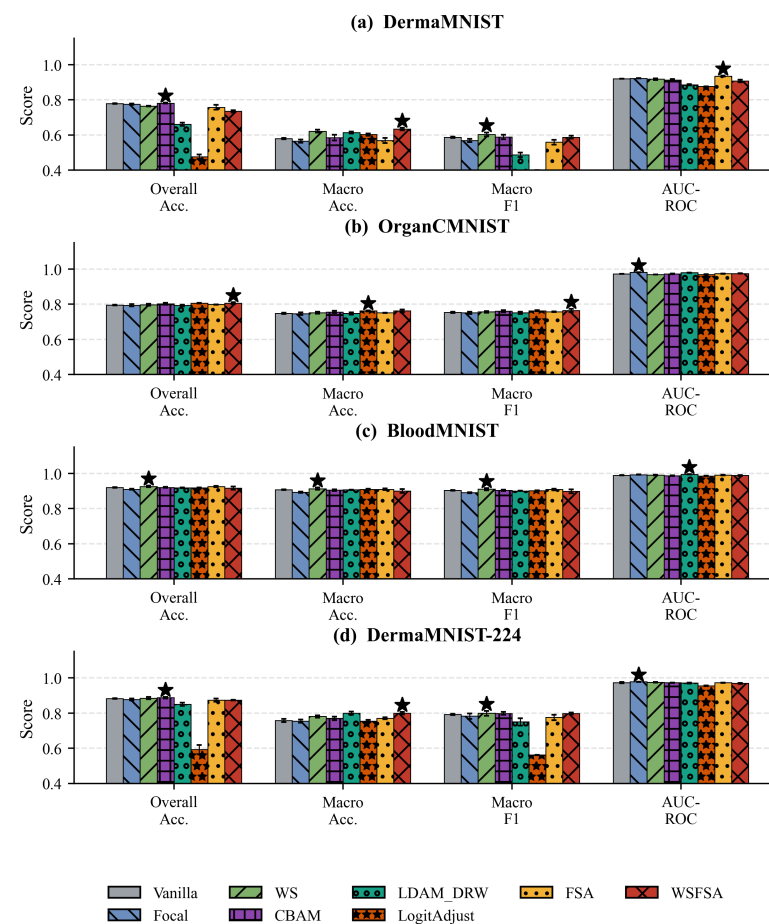
#### 3.2. Summary Metrics

Figure 5 and Table 1 present the overall test-set performance of all eight methods on DermaMNIST, BloodMNIST, OrganCMNIST, and DermaMNIST-224. On DermaMNIST, CBAM achieves the highest overall accuracy ( $0.781 \pm 0.003$ ), while WSFSA achieves the best macro accuracy ( $0.633 \pm 0.007$ ). This represents a +5.4 percentage point (pp) improvement over Vanilla in imbalance-aware accuracy metrics. FSA achieves the highest AUC-ROC ( $0.935 \pm 0.003$ ), suggesting strong discriminative ability across operating thresholds. The gap between overall accuracy and macro accuracy for Vanilla (0.779 vs. 0.579) indicates that standard training remains biased toward majority classes under severe imbalance.

On BloodMNIST, WS achieves the best performance across most metrics, with an overall accuracy of  $0.926 \pm 0.004$ , macro accuracy of  $0.914 \pm 0.006$ , and macro F1 of  $0.910 \pm 0.006$ , while LDAM-DRW achieves the highest AUC-ROC of  $0.995 \pm 0.000$ . The smaller performance differences among methods suggest that rebalancing strategies provide less benefit when class imbalance is relatively mild. FSA and WSFSA remain competitive, but neither surpasses the best-performing baseline on this dataset.



**Figure 4.** Training curves for all experiments.



**Figure 5.** Summary of the statistical metrics for all experiments. The black star marker above a bar denotes the best-performing method for that evaluation metric.

**Table 1.** Test-set performance across four datasets. Results are reported as mean  $\pm$  standard deviation. Best mean value per column within each dataset is shown in **bold**.

| (a) DermaMNIST     |                                     |                                     |                                     |                                     |
|--------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| Method             | Overall Acc.                        | Macro Acc.                          | Macro F1                            | AUC-ROC                             |
| Vanilla            | 0.779 $\pm$ 0.002                   | 0.579 $\pm$ 0.004                   | 0.586 $\pm$ 0.004                   | 0.920 $\pm$ 0.001                   |
| Focal              | 0.775 $\pm$ 0.004                   | 0.565 $\pm$ 0.009                   | 0.569 $\pm$ 0.009                   | 0.922 $\pm$ 0.001                   |
| WS                 | 0.764 $\pm$ 0.002                   | 0.621 $\pm$ 0.008                   | <b>0.603 <math>\pm</math> 0.012</b> | 0.917 $\pm$ 0.004                   |
| CBAM               | <b>0.781 <math>\pm</math> 0.003</b> | 0.585 $\pm$ 0.016                   | 0.588 $\pm$ 0.013                   | 0.912 $\pm$ 0.006                   |
| LDAM_DRW           | 0.661 $\pm$ 0.009                   | 0.614 $\pm$ 0.005                   | 0.487 $\pm$ 0.012                   | 0.886 $\pm$ 0.004                   |
| LogitAdjust        | 0.477 $\pm$ 0.013                   | 0.602 $\pm$ 0.007                   | 0.393 $\pm$ 0.007                   | 0.876 $\pm$ 0.001                   |
| FSA                | 0.757 $\pm$ 0.014                   | 0.568 $\pm$ 0.015                   | 0.559 $\pm$ 0.014                   | <b>0.935 <math>\pm</math> 0.003</b> |
| WSFSA              | 0.735 $\pm$ 0.006                   | <b>0.633 <math>\pm</math> 0.007</b> | 0.587 $\pm$ 0.008                   | 0.908 $\pm$ 0.007                   |
| (b) BloodMNIST     |                                     |                                     |                                     |                                     |
| Method             | Overall Acc.                        | Macro Acc.                          | Macro F1                            | AUC-ROC                             |
| Vanilla            | 0.921 $\pm$ 0.002                   | 0.907 $\pm$ 0.002                   | 0.904 $\pm$ 0.003                   | 0.990 $\pm$ 0.001                   |
| Focal              | 0.910 $\pm$ 0.003                   | 0.893 $\pm$ 0.003                   | 0.891 $\pm$ 0.003                   | 0.993 $\pm$ 0.000                   |
| WS                 | <b>0.926 <math>\pm</math> 0.004</b> | <b>0.914 <math>\pm</math> 0.006</b> | <b>0.910 <math>\pm</math> 0.006</b> | 0.991 $\pm$ 0.002                   |
| CBAM               | 0.921 $\pm$ 0.003                   | 0.905 $\pm$ 0.005                   | 0.904 $\pm$ 0.004                   | 0.988 $\pm$ 0.002                   |
| LDAM_DRW           | 0.918 $\pm$ 0.002                   | 0.907 $\pm$ 0.001                   | 0.901 $\pm$ 0.001                   | <b>0.995 <math>\pm</math> 0.000</b> |
| LogitAdjust        | 0.919 $\pm$ 0.002                   | 0.910 $\pm$ 0.002                   | 0.903 $\pm$ 0.002                   | 0.986 $\pm$ 0.001                   |
| FSA                | 0.926 $\pm$ 0.004                   | 0.910 $\pm$ 0.005                   | 0.909 $\pm$ 0.005                   | 0.991 $\pm$ 0.001                   |
| WSFSA              | 0.917 $\pm$ 0.009                   | 0.901 $\pm$ 0.011                   | 0.899 $\pm$ 0.011                   | 0.989 $\pm$ 0.002                   |
| (c) OrganCMNIST    |                                     |                                     |                                     |                                     |
| Method             | Overall Acc.                        | Macro Acc.                          | Macro F1                            | AUC-ROC                             |
| Vanilla            | 0.795 $\pm$ 0.001                   | 0.748 $\pm$ 0.004                   | 0.753 $\pm$ 0.003                   | 0.973 $\pm$ 0.001                   |
| Focal              | 0.794 $\pm$ 0.007                   | 0.746 $\pm$ 0.008                   | 0.751 $\pm$ 0.007                   | <b>0.981 <math>\pm</math> 0.000</b> |
| WS                 | 0.797 $\pm$ 0.006                   | 0.752 $\pm$ 0.006                   | 0.756 $\pm$ 0.005                   | 0.969 $\pm$ 0.001                   |
| CBAM               | 0.802 $\pm$ 0.006                   | 0.756 $\pm$ 0.007                   | 0.760 $\pm$ 0.007                   | 0.973 $\pm$ 0.001                   |
| LDAM_DRW           | 0.793 $\pm$ 0.005                   | 0.747 $\pm$ 0.007                   | 0.751 $\pm$ 0.007                   | 0.979 $\pm$ 0.001                   |
| LogitAdjust        | 0.805 $\pm$ 0.001                   | <b>0.762 <math>\pm</math> 0.003</b> | 0.763 $\pm$ 0.003                   | 0.969 $\pm$ 0.001                   |
| FSA                | 0.799 $\pm$ 0.001                   | 0.751 $\pm$ 0.001                   | 0.757 $\pm$ 0.002                   | 0.974 $\pm$ 0.001                   |
| WSFSA              | <b>0.806 <math>\pm</math> 0.005</b> | 0.762 $\pm$ 0.008                   | <b>0.765 <math>\pm</math> 0.009</b> | 0.975 $\pm$ 0.001                   |
| (d) DermaMNIST-224 |                                     |                                     |                                     |                                     |
| Method             | Overall Acc.                        | Macro Acc.                          | Macro F1                            | AUC-ROC                             |
| Vanilla            | 0.882 $\pm$ 0.002                   | 0.758 $\pm$ 0.010                   | 0.793 $\pm$ 0.003                   | 0.973 $\pm$ 0.003                   |
| Focal              | 0.877 $\pm$ 0.006                   | 0.753 $\pm$ 0.009                   | 0.782 $\pm$ 0.015                   | <b>0.978 <math>\pm</math> 0.001</b> |
| WS                 | 0.885 $\pm$ 0.006                   | 0.781 $\pm$ 0.006                   | <b>0.799 <math>\pm</math> 0.013</b> | 0.975 $\pm$ 0.002                   |
| CBAM               | <b>0.887 <math>\pm</math> 0.005</b> | 0.770 $\pm$ 0.009                   | 0.798 $\pm$ 0.008                   | 0.972 $\pm$ 0.001                   |
| LDAM_DRW           | 0.850 $\pm$ 0.010                   | 0.799 $\pm$ 0.009                   | 0.750 $\pm$ 0.021                   | 0.971 $\pm$ 0.003                   |
| LogitAdjust        | 0.593 $\pm$ 0.026                   | 0.754 $\pm$ 0.006                   | 0.562 $\pm$ 0.001                   | 0.956 $\pm$ 0.001                   |
| FSA                | 0.873 $\pm$ 0.009                   | 0.770 $\pm$ 0.006                   | 0.776 $\pm$ 0.015                   | 0.973 $\pm$ 0.001                   |
| WSFSA              | 0.874 $\pm$ 0.002                   | <b>0.800 <math>\pm</math> 0.007</b> | 0.797 $\pm$ 0.006                   | 0.969 $\pm$ 0.002                   |

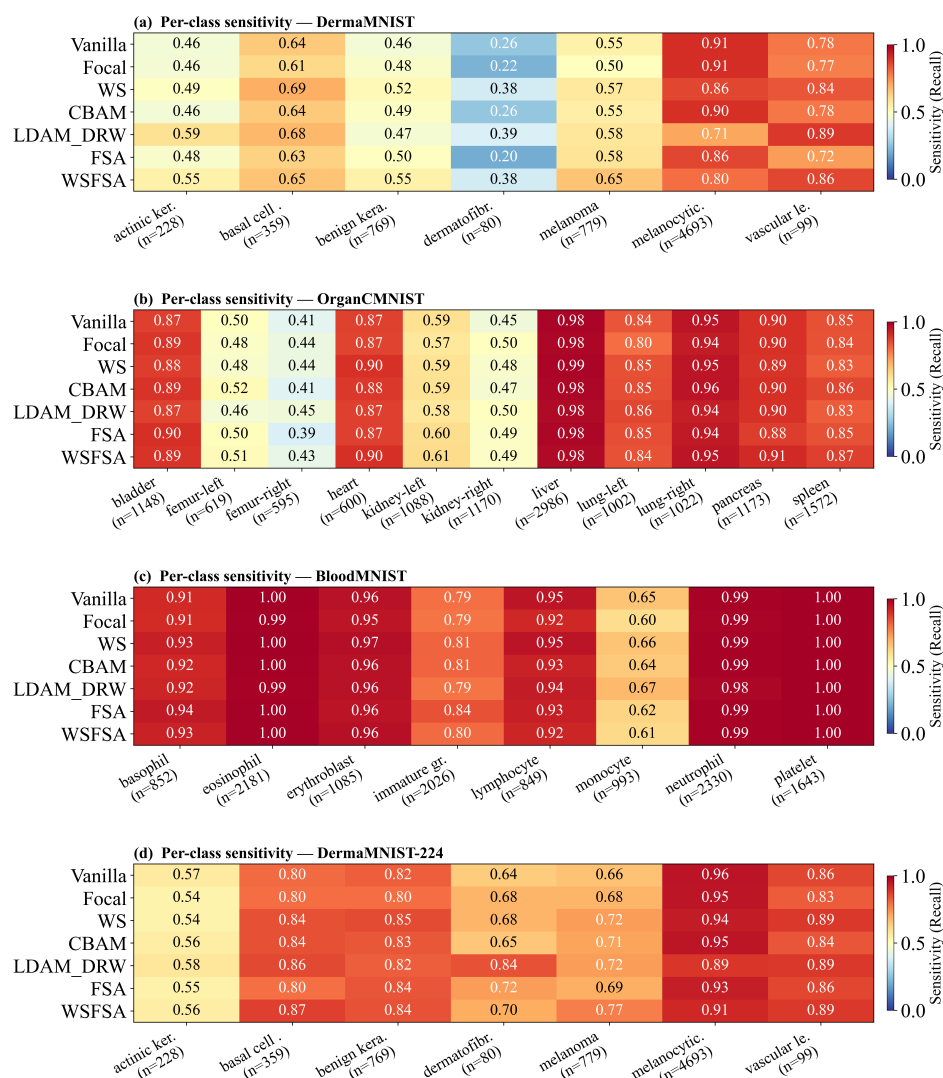
On OrganCMNIST, WSFSA achieves the best overall accuracy (0.806  $\pm$  0.005) and macro F1 (0.765  $\pm$  0.009), and LogitAdjust achieves the best macro accuracy (0.762  $\pm$  0.008). Compared with Vanilla, WSFSA improves overall accuracy by +1.1 pp, macro accuracy by +1.4 pp, and macro F1 by +1.2 pp. Focal loss achieves the highest AUC-ROC (0.981  $\pm$  0.000), but its macro accuracy and macro F1 are lower than those of WSFSA. Overall, these results show that WSFSA provides the clearest benefit on DermaMNIST and DermaMNIST-224, with moderate gains on OrganCMNIST, while remaining competitive on BloodMNIST.

On DermaMNIST-224, CBAM achieves the highest overall accuracy (0.887  $\pm$  0.005), while WSFSA achieves the best macro accuracy (0.800  $\pm$  0.007). Compared with Vanilla, WSFSA improves macro accuracy by +4.2 pp, showing that the proposed method remains effective under a higher-resolution setting for the most imbalanced benchmark. WS achieves

the highest macro F1 ( $0.799 \pm 0.013$ ), while Focal loss achieves the highest AUC-ROC ( $0.978 \pm 0.001$ ). Overall, the DermaMNIST-224 results further support that WSFSA is most beneficial for improving class-balanced performance under severe class imbalance.

### 3.2.1. Per-Class Sensitivity

Figure 6 shows the average per-class sensitivity (recall) of all methods on DermaMNIST, OrganCMNIST, BloodMNIST, and DermaMNIST-224. On DermaMNIST, WSFSA improves the sensitivity of several minority and clinically important classes. Compared with Vanilla, WSFSA increases sensitivity for actinic keratoses from 0.46 to 0.55, benign keratosis-like lesions from 0.46 to 0.55, dermatofibroma from 0.26 to 0.38, melanoma from 0.55 to 0.65, and vascular lesions from 0.78 to 0.86. WSFSA achieves the highest sensitivity for benign keratosis-like lesions and melanoma, while LDAM-DRW achieves the highest sensitivity for actinic keratoses, dermatofibroma, and vascular lesions. These results indicate that WSFSA improves recall for several minority classes, although the improvement is not uniform across all rare categories.



**Figure 6.** Per-class sensitivity (recall) of all experiments.

A similar pattern is observed on DermaMNIST-224, where WSFSA achieves the highest sensitivity for basal cell carcinoma and melanoma, and remains competitive for vascular lesions. However, other methods perform better for some classes, such as LDAM-DRW on

dermatofibroma. On OrganCMNIST and BloodMNIST, the per-class sensitivity differences are generally smaller, consistent with the milder class imbalance in these datasets.

On BloodMNIST, all methods achieve high sensitivity for most classes, reflecting the weaker imbalance effect in this dataset. The main difficult class is monocyte, for which LDAM-DRW achieves the highest sensitivity (0.67), followed by WS (0.66) and Vanilla (0.65). WSFSA remains competitive overall, although it does not improve monocyte sensitivity on this dataset. For several other classes, including eosinophil, neutrophil, and platelet, the sensitivities are close to saturation across methods, leaving limited room for improvement.

On OrganCMNIST, WSFSA provides competitive and often improved sensitivity across multiple organs. It achieves the highest sensitivity for kidney-left (0.61), pancreas (0.91), and spleen (0.87), and is tied for the highest sensitivity for heart (0.90). However, other methods perform better for some classes, such as FSA for bladder, CBAM for femur-left and lung-right, LDAM-DRW for femur-right and lung-left, and WS for liver. Overall, the per-class results show that WSFSA improves sensitivity for several minority or difficult classes, particularly on DermaMNIST and OrganCMNIST, while maintaining competitive recall on the less imbalanced BloodMNIST dataset.

A tradeoff is also observed on DermaMNIST: WSFSA reduces sensitivity for the majority class, melanocytic nevi, from 0.91 with Vanilla to 0.80. This reduction is expected for rebalancing-based methods, which shift the model away from majority-class dominance and toward improved recognition of minority classes.

### 3.2.2. Per-Class Specificity

Figure 7 presents the per-class specificity of all eight methods on DermaMNIST, OrganCMNIST, BloodMNIST, and DermaMNIST-224. Overall, specificity remains high across the four datasets, indicating that the sensitivity improvements observed in Figure 6 are not mainly caused by indiscriminate over-prediction of minority classes.

On DermaMNIST, most classes maintain specificity above 0.95 across methods. The main exception is the majority class, melanocytic nevi, where Vanilla has a lower specificity of 0.79, suggesting that many non-nevi samples are incorrectly predicted as nevi. WSFSA increases the specificity of this class to 0.90, indicating reduced false positives for the majority class. This improvement is accompanied by modest decreases in specificity for some minority classes, such as melanoma, where WSFSA obtains 0.88 compared with 0.93 for Vanilla.

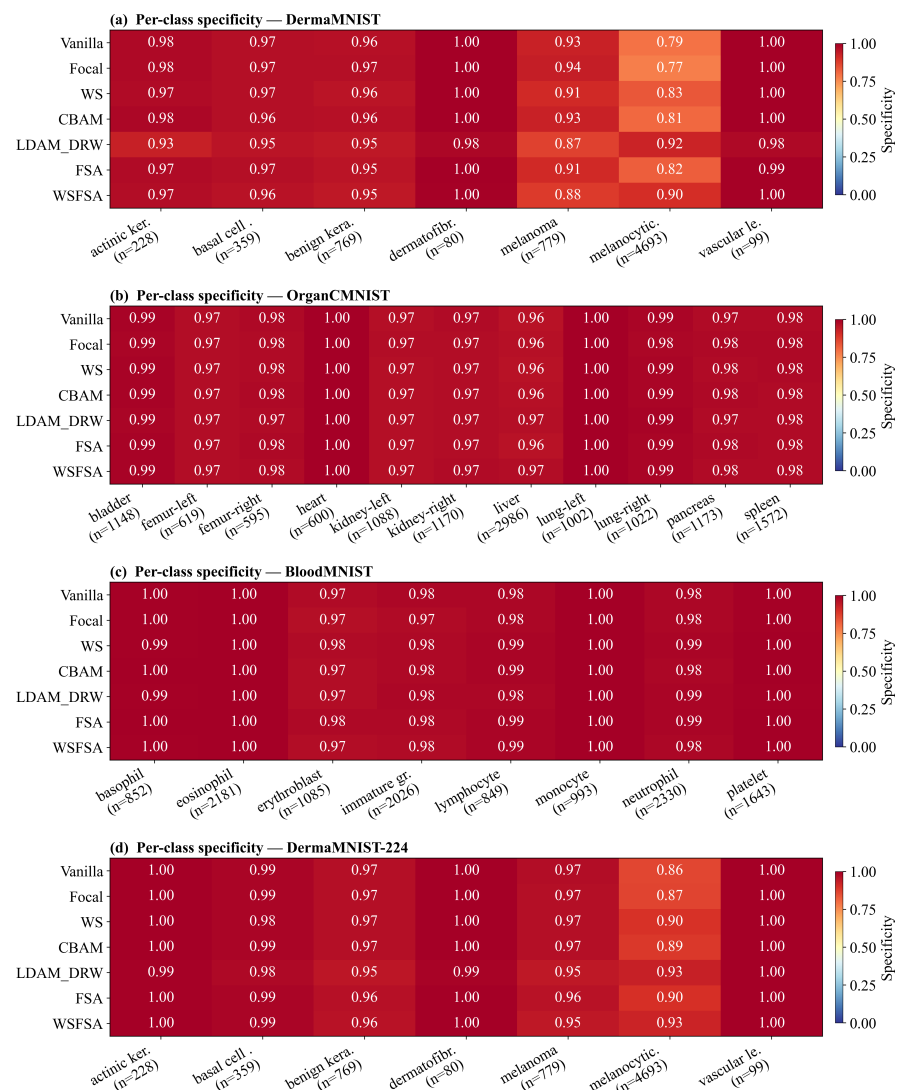
On DermaMNIST-224, a similar pattern is observed. Specificity remains high for most classes, while WSFSA improves the specificity of melanocytic nevi from 0.86 with Vanilla to 0.93. This suggests that the proposed method also reduces majority-class false positives in the higher-resolution setting.

On BloodMNIST and OrganCMNIST, all methods achieve consistently high specificity, mostly between 0.97 and 1.00. The differences among methods are small, suggesting that the rebalancing strategies do not substantially increase false-positive rates on the less severely imbalanced datasets. Overall, Figure 7 shows that WSFSA improves or maintains competitive sensitivity while preserving high specificity across most classes.

### 3.2.3. Joint Sensitivity and Specificity Analysis

Figure 8 presents a bubble matrix summarizing per-class sensitivity (bubble size) and specificity (color) across all four datasets, allowing simultaneous assessment of both metrics. On DermaMNIST, minority classes such as dermatofibroma ( $n = 80$ ), actinic keratoses ( $n = 228$ ), and vascular lesions ( $n = 99$ ) display smaller bubbles under Vanilla and Focal, reflecting weaker detection of rare conditions. WSFSA enlarges several minority-class bubbles, particularly for actinic keratoses, benign keratosis-like lesions, dermatofibroma,

melanoma, and vascular lesions. Meanwhile, the majority class, melanocytic nevi, shifts from lighter coloring under Vanilla to darker red under WSFSA, indicating reduced false positives for this dominant class.



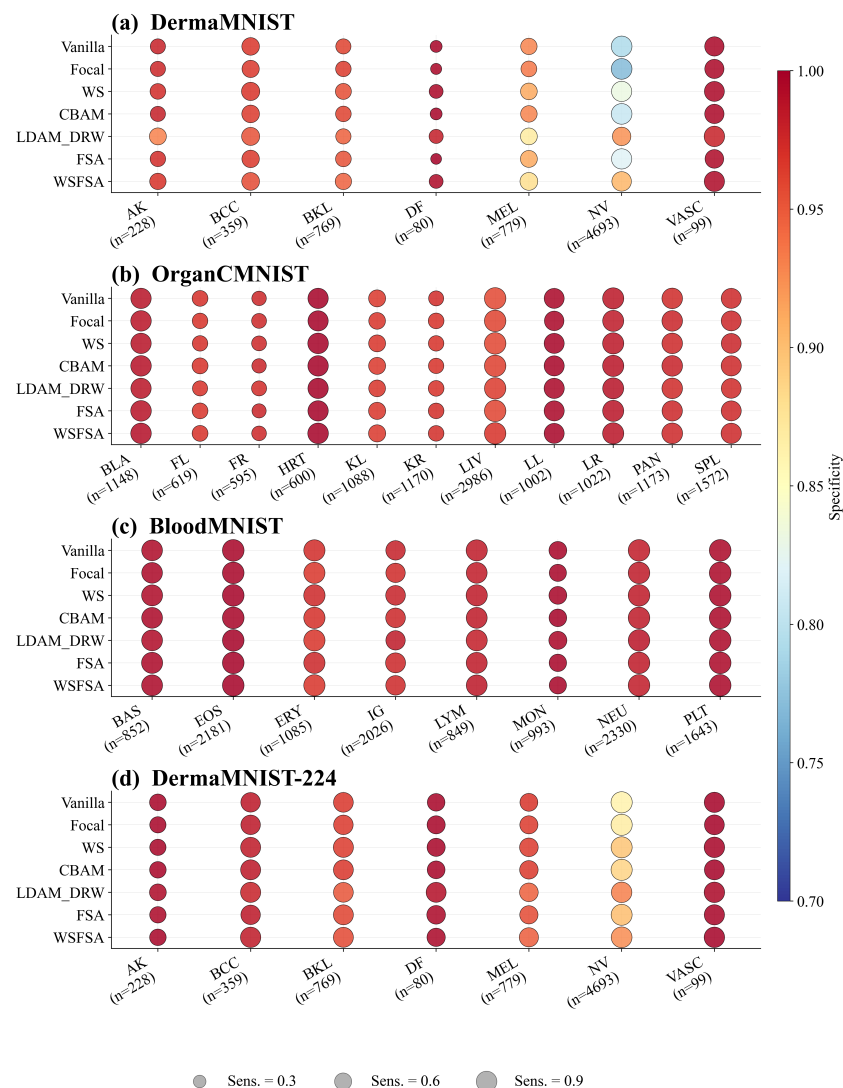
**Figure 7.** Per-class specificity across four datasets.

On DermaMNIST-224, a similar pattern is observed in the higher-resolution setting. WSFSA improves sensitivity for several minority or clinically important categories, including basal cell carcinoma and melanoma, while maintaining high specificity across most classes. The specificity of melanocytic nevi also improves under WSFSA, suggesting reduced majority-class false positives.

On BloodMNIST, bubble sizes are uniformly large across most methods and classes, reflecting the weaker imbalance effect in this dataset. The monocyte class remains the most difficult category, with a smaller bubble across methods. WSFSA remains competitive overall, although it does not provide the clearest improvement on this less imbalanced dataset. On OrganCMNIST, several anatomically similar classes, such as femur-left/femur-right and kidney-left/kidney-right, show relatively smaller bubbles across methods, likely due to visual similarity. WSFSA improves or remains competitive for several organs, including kidney-left, heart, pancreas, and spleen.

Overall, Figure 8 shows that WSFSA provides the clearest benefit on the severely imbalanced dermatology datasets, where it improves minority-class sensitivity while

preserving high specificity for most classes. The results also highlight an expected tradeoff: rebalancing improves recognition of several underrepresented classes but may reduce sensitivity for the majority class, such as melanocytic nevi.



**Figure 8.** Per-class sensitivity–specificity tradeoff across four benchmarks. Class abbreviations. DermaMNIST: AK, actinic keratoses and intraepithelial carcinoma; BCC, basal cell carcinoma; BKL, benign keratosis-like lesions; DF, dermatofibroma; MEL, melanoma; NV, melanocytic nevi; VASC, vascular lesions. OrganCMNIST: BLA, bladder; FL, femur-left; FR, femur-right; HRT, heart; KL, kidney-left; KR, kidney-right; LIV, liver; LL, lung-left; LR, lung-right; PAN, pancreas; SPL, spleen. BloodMNIST: BAS, basophil; EOS, eosinophil; ERY, erythroblast; IG, immature granulocytes; LYM, lymphocyte; MON, monocyte; NEU, neutrophil; PLT, platelet.

### 3.3. Ablation and Complexity Analysis

Table 2 reports an ablation on DermaMNIST that isolates the contribution of each component with three random seeds. Starting from the Vanilla baseline, WS improves macro accuracy from 0.579 to 0.622, indicating improved minority-class performance at the expected cost of overall accuracy. WSFSA further improves macro accuracy to 0.633, showing that FSA contributes beyond resampling alone. In contrast, applying FSA without weighted sampling does not improve macro accuracy over the baseline, suggesting that the two components are most effective in combination. Adding a class-prior logit adjustment to WSFSA reduces macro accuracy from 0.633 to 0.605, likely because weighted sampling and

FSA already address the imbalance and the additional correction leads to overcompensation. We therefore omit the logit bias from the final model.

To assess deployment feasibility, we measured the parameter count, FLOPs, and per-image inference latency of each model, as reported in Table 3. FSA increases the ResNet-18 backbone by only 1.18% in parameters, mainly due to the auxiliary pre-classifier head, with a negligible FLOP increase and a small latency increase from 0.1834 to 0.1877 ms per image on a Tesla T4. Because the module operates only on the final  $7 \times 7$  feature map, its computational overhead is minimal and does not substantially affect practical deployment.

We further analyze the error introduced when the inference-time attention weight,  $\omega = \sum_c \hat{p}(c | x)w_c$ , is estimated from the pre-classifier softmax output  $\hat{p}$ . In isolation, the pre-classifier is weak, with a test accuracy of 0.193 and an expected calibration error of 0.024 after temperature scaling. Its recall is also low for several minority classes, such as melanoma (0.013) and vascular lesions (0.034), indicating a bias toward majority classes. This bias propagates to  $\omega$ . Because  $\hat{p}$  tends to assign most probability mass to majority classes,  $\omega$  underestimates the target weight for minority-class inputs. Table 4 reports the per-class error,  $\delta_c = \mathbb{E}[\omega | y = c] - w_c$ . The error is strongly negative for the rarest classes, such as dermatofibroma (−0.477), and positive for the majority class, melanocytic nevi (+0.401). We also explored calibrating the pre-classifier outputs before computing  $\omega$  using a validation-fit prior correction, but the optimal correction was negligible because reducing minority-class error introduced larger majority-class error. We therefore treat calibration of the pre-classifier outputs as future work and note the dependence of  $\omega$  on pre-classifier quality as a limitation.

**Table 2.** Ablation on DermaMNIST isolating WS, FSA, and the FSA logit bias. Results are reported as mean  $\pm$  standard deviation over the evaluated seeds. Best mean value per column is shown in **bold**.

| Configuration       | Overall Acc.                        | Macro Acc.                          | Macro F1                            | AUC-ROC                             |
|---------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| Vanilla (baseline)  | <b>0.779 <math>\pm</math> 0.002</b> | 0.579 $\pm$ 0.005                   | 0.584 $\pm$ 0.003                   | 0.920 $\pm$ 0.000                   |
| WS                  | 0.763 $\pm$ 0.002                   | 0.622 $\pm$ 0.009                   | <b>0.601 <math>\pm</math> 0.014</b> | 0.918 $\pm$ 0.005                   |
| FSA (no logit bias) | 0.769 $\pm$ 0.003                   | 0.589 $\pm$ 0.005                   | 0.581 $\pm$ 0.004                   | <b>0.935 <math>\pm</math> 0.003</b> |
| FSA                 | 0.749 $\pm$ 0.009                   | 0.559 $\pm$ 0.009                   | 0.549 $\pm$ 0.000                   | 0.933 $\pm$ 0.000                   |
| WSFSA               | 0.732 $\pm$ 0.006                   | <b>0.633 <math>\pm</math> 0.007</b> | 0.590 $\pm$ 0.009                   | 0.904 $\pm$ 0.006                   |
| WSFSA + logit bias  | 0.734 $\pm$ 0.019                   | 0.605 $\pm$ 0.016                   | 0.566 $\pm$ 0.002                   | 0.905 $\pm$ 0.010                   |

**Table 3.** Model complexity and inference efficiency comparison.

| Model               | Params     | Params (M) | GFLOPs | Latency (ms) | Std.   | Batch Size | Image Size | Hardware |
|---------------------|------------|------------|--------|--------------|--------|------------|------------|----------|
| ResNet-18 (Vanilla) | 11,180,103 | 11.1801    | 0.6699 | 0.1834       | 0.0013 | 64         | 96         | Tesla T4 |
| ResNet-18 + CBAM    | 11,213,514 | 11.2135    | 0.6700 | 0.1863       | 0.0013 | 64         | 96         | Tesla T4 |
| ResNet-18 + FSA     | 11,311,794 | 11.3118    | 0.6702 | 0.1877       | 0.0010 | 64         | 96         | Tesla T4 |

**Table 4.** Per-class approximation error of the inference-time attention weight on DermaMNIST.  $w_c$  denotes the max-normalized inverse-frequency target weight, and  $\delta_c = \mathbb{E}[\omega | y = c] - w_c$  denotes the estimation error.

| Class                                         | Train $n$ | $w_c$ | $\delta_c$ |
|-----------------------------------------------|-----------|-------|------------|
| Dermatofibroma                                | 80        | 1.000 | −0.477     |
| Vascular lesions                              | 99        | 0.899 | −0.367     |
| Actinic keratoses & intraepithelial carcinoma | 228       | 0.592 | −0.045     |
| Basal cell carcinoma                          | 359       | 0.472 | +0.069     |
| Benign keratosis-like lesions                 | 769       | 0.323 | +0.200     |
| Melanoma                                      | 779       | 0.320 | +0.199     |
| Melanocytic nevi                              | 4693      | 0.131 | +0.401     |

## 4. Discussion

The results show that WSFSA can improve sensitivity for minority classes without causing a large drop in specificity. This is because WS and FSA help the model in different ways. WS changes which samples the model sees during training, while FSA changes how the model attends to features from different classes. When used together, they give better balanced performance than using either method alone.

There are still two main limitations. First, although we tested the method on four MedMNIST datasets with different imbalance levels, the results are still based on benchmark datasets. Therefore, further validation on larger and more diverse clinical datasets is needed. Second, several settings, including the frequency-weighting exponent  $\beta$  and the logit bias, are selected manually. In particular,  $\beta$  is implemented as a single global value shared across all classes. The relatively low sensitivity of FSA on dermatofibroma suggests that a fixed global  $\beta$  may be insufficient when classes differ substantially in rarity or visual heterogeneity, because it cannot adapt the strength of frequency weighting on a per-class basis.

Future work can improve the method in multiple ways. First, WSFSA could be tested on more real-world datasets with different levels of class imbalance to better understand when it works best. Second, the fixed global frequency-weighting scheme could be replaced with a learnable or class-dependent strategy. For example, the global exponent  $\beta$  could be extended to a per-class parameter  $\beta_c$ , either learned during training or estimated from class statistics. Alternatively, temperature scaling could be applied to the inference-time pre-classifier before computing the expected frequency weight. These extensions may reduce manual tuning and improve robustness for extremely rare or visually heterogeneous classes. Moreover, a more rigorous statistical significance analysis based on a larger number of independent runs would further strengthen the work.

## 5. Conclusions

In this study, we proposed WSFSA, a frequency-aware spatial attention module combined with weighted sampling for class-imbalanced image classification. The method addresses imbalance from two complementary perspectives: WS increases the training exposure of minority classes, while FSA adaptively enhances spatial feature responses for underrepresented classes. Experiments on DermaMNIST, BloodMNIST, OrganCMNIST, and DermaMNIST-224 demonstrate that the effectiveness of WSFSA depends on the severity of class imbalance. On DermaMNIST, where imbalance is most severe, WSFSA achieved the best macro accuracy ( $0.633 \pm 0.007$ ), improving over the Vanilla baseline by 5.4 percentage points. On DermaMNIST-224, WSFSA achieved a macro accuracy of  $0.800 \pm 0.007$ , performing competitively with the strongest baseline methods and further demonstrating its effectiveness under a higher-resolution, severely imbalanced setting. On OrganCMNIST, WSFSA achieved the best overall accuracy and macro F1-score, and tied for the best macro accuracy. On BloodMNIST, where class imbalance is relatively mild, all methods performed similarly, and WS achieved the strongest overall class-balanced performance.

The per-class analysis further shows that WSFSA improves sensitivity for several minority or difficult classes, such as actinic keratoses, dermatofibroma, melanoma, and vascular lesions in DermaMNIST, as well as kidney-left, pancreas, and spleen in OrganCMNIST. These gains are achieved while maintaining generally high specificity across classes, indicating that WSFSA does not simply over-predict minority classes. Overall, the results suggest that combining sampling-level and feature-level rebalancing is a practical strategy for improving class-balanced recognition, particularly when the dataset exhibits severe class imbalance.

**Author Contributions:** Conceptualization, P.L.; methodology, S.Z. and P.L.; software, S.Z.; validation, S.Z.; formal analysis, S.Z.; investigation, S.Z.; writing—original draft preparation, S.Z.; writing—review and editing, P.L.; supervision, P.L. All authors have read and agreed to the published version of the manuscript.

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

**Institutional Review Board Statement:** Ethical review and approval were waived for this study due to this study uses the publicly available benchmark dataset from the MedMNIST v2 collection, which contains fully anonymised images. No human subjects were recruited and no identifiable patient data were accessed.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are openly available in the MedMNIST v2 collection at <https://medmnist.com> (accessed on 28 June 2026).

**Acknowledgments:** During the preparation of this manuscript, the authors used Claude Opus 4.8 for the purposes of code debugging and ChatGPT 5.5 for the purpose of text polishing. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

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

## Abbreviations

The following abbreviations are used in this manuscript:

|       |                                                          |
|-------|----------------------------------------------------------|
| DNN   | Deep neural network                                      |
| FSA   | Frequency-aware spatial attention                        |
| pp    | Percentage point                                         |
| WS    | Weighted sampling                                        |
| WSFSA | Weighted sampling with frequency-aware spatial attention |

## References

1. Benenson, R.; Popov, S.; Ferrari, V. Large-scale interactive object segmentation with human annotators. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2019; pp. 11700–11709.
2. Wang, Y.; Ahsan, U.; Li, H.; Hagen, M. A comprehensive review of modern object segmentation approaches. *Found. Trends Comput. Graph. Vis.* **2022**, *13*, 111–283. [[CrossRef](#)]
3. Gao, M.; Zheng, F.; Yu, J.J.; Shan, C.; Ding, G.; Han, J. Deep learning for video object segmentation: A review. *Artif. Intell. Rev.* **2023**, *56*, 457–531.
4. Chandola, V.; Banerjee, A.; Kumar, V. Anomaly detection: A survey. *ACM Comput. Surv. (CSUR)* **2009**, *41*, 1–58. [[CrossRef](#)]
5. Bhuyan, M.H.; Bhattacharyya, D.K.; Kalita, J.K. Network anomaly detection: Methods, systems and tools. *IEEE Commun. Surv. Tutor.* **2013**, *16*, 303–336. [[CrossRef](#)]
6. Tang, M.; Liu, M.; Li, H.; Yang, J.; Wei, C.; Li, B.; Li, D.; Xu, R.; Xu, Y.; Zhang, Z.; et al. Async learned user embeddings for ads delivery optimization. *arXiv* **2024**, arXiv:2406.05898.
7. Hsu, Y.C.; Lv, Z.; Schlosser, J.; Odom, P.; Kira, Z. Multi-class classification without multi-class labels. *arXiv* **2019**, arXiv:1901.00544.
8. Grandini, M.; Bagli, E.; Visani, G. Metrics for multi-class classification: An overview. *arXiv* **2020**, arXiv:2008.05756.
9. Xie, Y.; Chang, H.; Li, Z.; Liang, L.; Chen, X.; Zhao, D. A unified framework for locating and recognizing human actions. In *Proceedings of the CVPR 2011*; IEEE: New York, NY, USA, 2011; pp. 25–32.
10. Liu, N.; Li, X.; Qi, E.; Xu, M.; Li, L.; Gao, B. A novel ensemble learning paradigm for medical diagnosis with imbalanced data. *IEEE Access* **2020**, *8*, 171263–171280. [[CrossRef](#)]
11. Huynh, T.; Nibali, A.; He, Z. Semi-supervised learning for medical image classification using imbalanced training data. *Comput. Methods Programs Biomed.* **2022**, *216*, 106628. [[CrossRef](#)] [[PubMed](#)]
12. Ngan, H.Y.; Pang, G.K.; Yung, N.H. Automated fabric defect detection—A review. *Image Vis. Comput.* **2011**, *29*, 442–458. [[CrossRef](#)]
13. Ren, Z.; Fang, F.; Yan, N.; Wu, Y. State of the art in defect detection based on machine vision. *Int. J. Precis. Eng. Manuf.-Green Technol.* **2022**, *9*, 661–691.
14. Majjala, P.; Shuyang, Z.; Heittola, T.; Virtanen, T. Environmental noise monitoring using source classification in sensors. *Appl. Acoust.* **2018**, *129*, 258–267. [[CrossRef](#)]

15. Singh, J.P.; Dwivedi, Y.K.; Rana, N.P.; Kumar, A.; Kapoor, K.K. Event classification and location prediction from tweets during disasters. *Ann. Oper. Res.* **2019**, *283*, 737–757.
16. Gill, K.S.; Anand, V.; Upadhyay, D.; Dangi, S. Classification of tweets using a machine learning and natural language processing algorithm for disaster prediction. In *Proceedings of the 2024 3rd International Conference for Innovation in Technology (INOCON)*; IEEE: New York, NY, USA, 2024; pp. 1–5.
17. Barua, S.; Islam, M.M.; Yao, X.; Murase, K. MWMOTE—majority weighted minority oversampling technique for imbalanced data set learning. *IEEE Trans. Knowl. Data Eng.* **2012**, *26*, 405–425.
18. Castellanos, F.J.; Valero-Mas, J.J.; Calvo-Zaragoza, J.; Rico-Juan, J.R. Oversampling imbalanced data in the string space. *Pattern Recognit. Lett.* **2018**, *103*, 32–38. [\[CrossRef\]](#)
19. Anand, A.; Pugalenth, G.; Fogel, G.B.; Suganthan, P. An approach for classification of highly imbalanced data using weighting and undersampling. *Amino Acids* **2010**, *39*, 1385–1391. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Lin, W.C.; Tsai, C.F.; Hu, Y.H.; Jhang, J.S. Clustering-based undersampling in class-imbalanced data. *Inf. Sci.* **2017**, *409*, 17–26.
21. Zhou, Z.; Zheng, C.; Liu, X.; Tian, Y.; Chen, X.; Chen, X.; Dong, Z. A dynamic effective class balanced approach for remote sensing imagery semantic segmentation of imbalanced data. *Remote Sens.* **2023**, *15*, 1768. [\[CrossRef\]](#)
22. Akil, M.; Saouli, R.; Kachouri, R. Fully automatic brain tumor segmentation with deep learning-based selective attention using overlapping patches and multi-class weighted cross-entropy. *Med. Image Anal.* **2020**, *63*, 101692.
23. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision*; IEEE: New York, NY, USA, 2017; pp. 2980–2988.
24. Mulyanto, M.; Faisal, M.; Prakosa, S.W.; Leu, J.S. Effectiveness of focal loss for minority classification in network intrusion detection systems. *Symmetry* **2020**, *13*, 4. [\[CrossRef\]](#)
25. Perry, T.; Bader-El-Den, M.; Cooper, S. Imbalanced classification using genetically optimized cost sensitive classifiers. In *Proceedings of the 2015 IEEE Congress on Evolutionary Computation (CEC)*; IEEE: New York, NY, USA, 2015; pp. 680–687.
26. Zhang, C.; Tan, K.C.; Li, H.; Hong, G.S. A cost-sensitive deep belief network for imbalanced classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *30*, 109–122. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Xu, Z.; Lee, C.K.M.; Lv, Y.; Chan, J. Ensemble capsule network with an attention mechanism for the fault diagnosis of bearings from imbalanced data samples. *Sensors* **2022**, *22*, 5543. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Wang, L.; Zhang, L.; Qi, X.; Yi, Z. Deep attention-based imbalanced image classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 3320–3330. [\[CrossRef\]](#)
29. Oubara, A.; Wu, F.; Maleki, R.; Ma, B.; Amamra, A.; Yang, G. Enhancing adversarial learning-based change detection in imbalanced datasets using artificial image generation and attention mechanism. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 125. [\[CrossRef\]](#)
30. Zhang, K.; Wu, Z.; Yuan, D.; Luan, J.; Jia, J.; Meng, H.; Song, B. Re-Weighted Interval Loss for Handling Data Imbalance Problem of End-to-End Keyword Spotting. In *Proceedings of the Interspeech*; ISCA: Singapore, 2020; Volume 108, pp. 2567–2571.
31. Cui, J.; Zhong, Z.; Liu, S.; Yu, B.; Jia, J. Parametric contrastive learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*; IEEE: New York, NY, USA, 2021; pp. 715–724.
32. Cong, C.; Liu, S.; Rana, P.; Pagnucco, M.; Di Ieva, A.; Berkovsky, S.; Song, Y. Adaptive unified contrastive learning with graph-based feature aggregator for imbalanced medical image classification. *Expert Syst. Appl.* **2024**, *251*, 123783. [\[CrossRef\]](#)
33. Marrakchi, Y.; Makansi, O.; Brox, T. Fighting class imbalance with contrastive learning. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 466–476.
34. Blagus, R.; Lusa, L. SMOTE for high-dimensional class-imbalanced data. *BMC Bioinform.* **2013**, *14*, 106.
35. Bhagat, R.C.; Patil, S.S. Enhanced SMOTE algorithm for classification of imbalanced big-data using random forest. In *Proceedings of the 2015 IEEE International Advance Computing Conference (IACC)*; IEEE: New York, NY, USA, 2015; pp. 403–408.
36. Yin, L.; Ge, Y.; Xiao, K.; Wang, X.; Quan, X. Feature selection for high-dimensional imbalanced data. *Neurocomputing* **2013**, *105*, 3–11. [\[CrossRef\]](#)
37. Suh, S.; Lukowicz, P.; Lee, Y.O. Discriminative feature generation for classification of imbalanced data. *Pattern Recognit.* **2022**, *122*, 108302. [\[CrossRef\]](#)
38. Fan, Y.; Shen, D.; Gur, R.C.; Gur, R.E.; Davatzikos, C. COMPARE: Classification of morphological patterns using adaptive regional elements. *IEEE Trans. Med. Imaging* **2007**, *26*, 93–105. [\[PubMed\]](#)
39. Quéllec, G.; Laniard, M.; Cazuguel, G.; Abràmoff, M.D.; Cochener, B.; Roux, C. Weakly supervised classification of medical images. In *Proceedings of the 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*; IEEE: New York, NY, USA, 2012; pp. 110–113.
40. Tong, W.; Chen, W.; Han, W.; Li, X.; Wang, L. Channel-attention-based DenseNet network for remote sensing image scene classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 4121–4132.
41. Hou, Y.E.; Yang, K.; Dang, L.; Liu, Y. Contextual spatial-channel attention network for remote sensing scene classification. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 6008805. [\[CrossRef\]](#)

42. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018*; IEEE: New York, NY, USA, 2018.
43. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)*; IEEE: New York, NY, USA, 2018; pp. 3–19.
44. Islam, O.; Kumer, K.; Akter, S.; Uddin, M.M. Multi-head self-attention mechanisms in vision transformers for retinal image classification. In *Proceedings of the 2024 IEEE International Conference on Computing, Applications and Systems (COMPAS)*; IEEE: New York, NY, USA, 2024; pp. 1–5.
45. Chen, Z.; Qi, Z.; Li, X.; Wang, Y.; Meng, L.; Meng, X. Class-aware convolution and attentive aggregation for image classification. In *Proceedings of the 5th ACM International Conference on Multimedia in Asia*; ACM Digital Library: New York, NY, USA, 2023; pp. 1–7.
46. Krasanakis, E.; Spyromitros-Xioufis, E.; Papadopoulos, S.; Kompatsiaris, Y. Adaptive sensitive reweighting to mitigate bias in fairness-aware classification. In *Proceedings of the 2018 World Wide Web Conference*; ACM Digital Library: New York, NY, USA, 2018; pp. 853–862.
47. Seiffert, C.; Khoshgoftaar, T.M.; Van Hulse, J.; Napolitano, A. RUSBoost: A hybrid approach to alleviating class imbalance. *IEEE Trans. Syst. Man. Cybern.-Part A Syst. Hum.* **2009**, *40*, 185–197. [[CrossRef](#)]
48. Yang, K.; Yu, Z.; Wen, X.; Cao, W.; Chen, C.P.; Wong, H.S.; You, J. Hybrid classifier ensemble for imbalanced data. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *31*, 1387–1400. [[CrossRef](#)] [[PubMed](#)]
49. Yang, J.; Shi, R.; Wei, D.; Liu, Z.; Zhao, L.; Ke, B.; Pfister, H.; Ni, B. Medmnist v2-a large-scale lightweight benchmark for 2d and 3d biomedical image classification. *Sci. Data* **2023**, *10*, 41. [[PubMed](#)]
50. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016*; IEEE: New York, NY, USA, 2016.

**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.