Mathematical Biosciences and Engineering

MBE, 22 (12): 3005–3027.
DOI: 10.3934/mbe.2025111
Received: 07 April 2025
Revised: 08 September 2025
Accepted: 25 September 2025
Published:09 October 2025

https://www.aimspress.com/journal/MBE

Research article

Few-shot learning for rare skin disease classification via adaptive distribution calibration

Yin Wen[1,2], Yingbo Wu[3], Zhigao Zeng[1,2,*], Shengqiu Yi[1,2], XinpanYuan[1,2] and Yanhui Zhu[1,2]

[1] School of Computer, Hunan University of Technology, Zhuzhou 412007, China
[2] Hunan Key Laboratory of Intelligent Information Perception and Processing Technology, Zhuzhou 412007, China
[3] Hubei Central China Technology Development of Electric Power Co., Ltd., Wuhan 430079, China

* **Correspondence:** Email: zzgzzg99@163.com.

Abstract: The classification of rare skin diseases faces significant data scarcity challenges due to the difficulty in acquiring clinical samples and the high cost of annotation, which severely hinders the training of deep neural network–based models. Few-shot learning has emerged as a cutting-edge solution, with its core capability being the identification of novel disease classes using limited annotated samples to mitigate data insufficiency. However, most existing methods fail to fully leverage the statistical information from base classes to calibrate the distribution of few-shot classes, thereby optimizing classifier inputs. Two critical research challenges remain: (1) accurately estimating the true distribution of few-shot classes with minimal samples, and (2) selecting appropriate base class information for effective distribution calibration. To address these challenges, we propose SADC (skin disease classification via adaptive distribution calibration), a new few-shot learning framework incorporating multi-scale feature extraction and adaptive sample calibration. First, our multi-scale feature extraction strategy employs feature descriptor matrices and composite metrics to optimize multi-dimensional, multi-directional feature representations, enabling precise similarity computation between base-class and few-shot samples. Second, the adaptive sample calibration strategy constructs weight matrices based on sample similarity to automatically select optimal base-class samples with adaptive weights for distribution calibration, ensuring alignment between calibrated distributions and true unbiased distributions. Experimental results demonstrated that SADC achieves state-of-the-art performance across three public dermatology datasets (ISIC2018, Derm7pt, and SD198), showing significant improvements over existing methods. The framework's innovation lies in its dual-strategy approach to distribution-aware few-shot learning, advancing the frontier of data-efficient medical

image analysis.

Keywords: Few-shot learning; Distribution calibration; Rare skin diseases; Multi-scale feature extraction; Adaptive sample calibration

1. Introduction

Skin diseases represent a global health concern with profound impacts on patients' quality of life, where early and accurate diagnosis is critical for effective treatment and improved prognosis [1]. Recent advances in deep learning have demonstrated significant potential in dermatology through medical image analysis [2–4]. However, privacy concerns and high annotation costs severely limit data accessibility, particularly for rare skin diseases that exhibit sparse samples typically residing in the long-tail region of data distributions (Figure 1). This data scarcity demands models with strong generalization capabilities to rapidly adapt and accurately classify novel disease categories under limited samples, making rare skin disease classification an unresolved challenge.

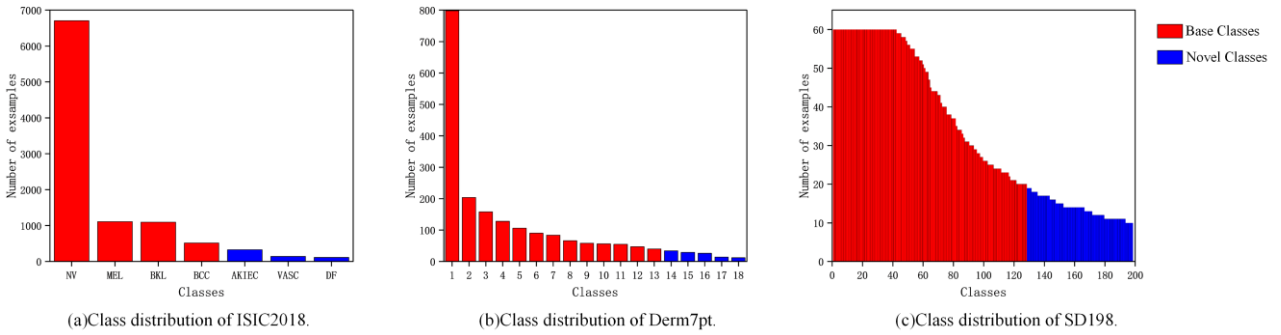

Figure 1. Class distribution of three datasets.

Few-shot learning (FSL) has emerged as a promising solution [5–7], with mainstream approaches falling into two categories: meta-learning and transfer learning. Meta-learning trains models to "learn how to learn", where model-agnostic meta-learning (MAML) [8] and prototypical networks (ProtoNet) [9] have been widely adopted and refined [10–15]. Transfer learning [16–20] leverages knowledge from related tasks or datasets by pretraining on base categories with sufficient samples, enabling quick adaptation to rare disease categories. Additionally, some studies exploit intra-class substructures to achieve finer classification in complex scenarios [20,22].

Despite progress, existing methods often underutilize base categories to calibrate few-shot class distributions for classifier optimization. Research indicates [23,24] that when features follow Gaussian distributions, similar categories share comparable statistics (e.g., mean and covariance). Transferring accurate statistics from analogous base categories can thus rectify biased distributions in few-shot classes. However, two critical challenges persist: (1) estimating true few-shot class distributions from minimal samples, and (2) selecting optimal base-class information for calibration.

To address these, we propose SADC (skin disease classification via adaptive distribution calibration), integrating:

1). The multi-scale feature extraction strategy that extracts multidimensional and multidirectional features to enhance similarity computation between base and novel classes;
2). The adaptive sample calibration strategy that dynamically selects base-class samples using similarity-weighted adjustments to align feature distributions with ground-truth representations.
3). Extensive experiments on three public dermatology datasets (ISIC2018, Derm7pt, and SD198) to demonstrate that the proposed method achieves state-of-the-art performance.

2. Related work

2.1. Few-shot learning

Few-shot learning (FSL) has emerged as a critical research direction in machine learning, addressing the challenge of enabling models to rapidly learn and generalize to novel categories with limited labeled samples [5–7]. This paradigm focuses on extracting prior knowledge from base classes (categories with abundant labeled samples) and effectively transferring it to novel classes (previously unseen categories with scarce annotations), thereby achieving accurate recognition even with minimal novel-class examples. Current FSL methodologies can be broadly categorized into two frameworks: meta-learning and transfer learning approaches.

2.1.1. Meta-learning-based methods

Inspired by meta-learning principles, these approaches have been widely adopted in FSL. Foundational models, including MAML [8], Reptile [25], Matching Networks [26], Prototypical Networks [9], and Relation Networks [27], employ meta-learning frameworks that train models in episodic batches. These methods enable networks to adapt parameters efficiently with few samples, achieving competitive accuracy on novel categories. Meta-learning approaches further divide into:

a). *Parameter optimization-based methods* (e.g., MAML, Reptile) that learn optimal initial parameters for rapid adaptation to unseen classes within a few gradient steps.
b). *Metric-based methods* that leverage encoded feature vectors with distance metrics (e.g., cosine similarity in Matching Networks, Euclidean distance in Prototypical Networks) for nearest-neighbor classification.

2.1.2. Transfer learning–based methods

Recently, transfer learning approaches have gained prominence in FSL. These methods first pretrain standard classification networks on base classes and then fine-tune classifier heads using few-shot batches from novel categories. By focusing on learning transferable feature representations—crucial for deep learning tasks—such approaches demonstrate remarkable effectiveness. Studies [28-29] reveal that even simple transfer learning baselines can match or surpass traditional FSL methods while offering simpler training pipelines. Given their demonstrated efficacy, we employ transfer learning as our foundational framework for rare skin disease prediction.

2.2. Few-shot learning for rare skin disease classification

In recent years, deep learning–based dermatological diagnostic systems have made significant strides in medical imaging analysis, demonstrating remarkable capabilities in skin disease classification. A seminal study by Gamage et al. [47] introduced an explainable melanoma classification framework that achieved 90.24% accuracy on the HAM10000 dataset through innovative modifications to the Xception architecture, incorporating Bayesian hyperparameter optimization and a strategic batch normalization layer unfreezing approach. Building upon this foundation, the research team further advanced the field [48] by developing a sophisticated hybrid CNN-ViT system that combines U2-Net segmentation modules with a novel Saliency Mask-Guided Vision Transformer architecture, pushing classification performance to an impressive 98.37% accuracy while maintaining clinical interpretability.

The classification of rare skin diseases presents significant challenges due to imbalanced category distributions and scarce image data. To address this issue, few-shot learning methods have been introduced to reduce reliance on large-scale training data while improving classification performance under limited data conditions.

In the parameter optimization sub-branch of meta-learning approaches, Mahajan et al. [10] enhanced model robustness to image variations by designing architectures that learn transformation-invariant features of dermatological images. Li et al. [11] incorporated a task difficulty awareness mechanism that dynamically adjusts task weights to improve model adaptation efficiency to novel classes. Furthermore, Singh et al. [12] strengthened model generalization through integration of advanced data augmentation techniques. In the metric-based sub-branch of meta-learning, Zhu et al. [13] addressed limitations of traditional cross-entropy loss for few-shot training scenarios by proposing a query-relative loss mechanism that effectively captures cross-sample correlations. Chowdhury et al. [14] achieved precise sample distribution modeling through maximum mean discrepancy (MMD)-weighted prototype networks, while Zhou et al. [15] improved classification performance by constructing adaptive subspaces incorporating multi-scale similarity metrics.

Under transfer learning approaches, Xiao et al. [16] developed a multi-task framework that enhances few-shot classification by incorporating contrastive learning as an auxiliary task. Dai et al. [17] proposed a dual-encoder architecture that combines knowledge from large-scale image datasets and few-shot medical image datasets. Other studies [18-20] have utilized unlabeled data and self-supervised learning to mitigate performance degradation in few-shot learners.

Additionally, research has focused on addressing intra-class variations in skin diseases. PCN [20] and SCAN [22] demonstrated that the same disease may exhibit significant appearance differences across body regions, achieving finer-grained classification in complex scenarios by discovering sub-cluster structures within categories.

Notably, in the field of rare skin disease classification, SS-DCN [30] remains one of the few studies employing distribution calibration to improve performance. SS-DCN calibrates novel class distributions by transferring statistical features (e.g., mean and covariance) from the most similar k base classes, leveraging similarity relationships between base and novel classes to achieve more accurate distribution estimation in few-shot scenarios. However, SS-DCN has a critical limitation: it focuses solely on determining what to transfer (which distribution information to extract from base classes) without optimizing how much to transfer (the contribution weights of different base classes to novel class calibration). Specifically, SS-DCN's equal weighting assumption considers all base classes

equally important, which overestimates distribution information from less similar base classes while underestimating more similar ones, potentially generating feature representations that deviate from true distributions.

To address this limitation, we propose SADC, a rare skin disease classification method based on adaptive distribution calibration. Our approach introduces two core methodological innovations: (1) a multi-scale feature extraction strategy that captures inter-dimensional relationships through cross-product operations to construct richer feature representations, and (2) an adaptive sample calibration mechanism that employs dynamic weight matrices and k-value selection to identify and leverage only the most relevant base class samples. These innovations enable three fundamental improvements: multi-scale representation overcomes the limitations of Euclidean distance in capturing complex distributions; adaptive calibration ensures utilization of semantically relevant base class information; and distribution calibration synthesizes more robust and representative feature distributions.

3. Methodology

3.1. Problem formulation

Given a standard labeled dataset $D = \{(x_i, y_i)\}$, where x_i denotes the i-th image, and y_i represents its corresponding class label, the dataset D is partitioned into a base class dataset D_b and a novel class dataset D_n, with the novel class containing only a limited number of samples. All classes C are divided into base classes C_b and novel classes C_n, satisfying $C_b \cup C_n = C, C_b \cap C_n = \emptyset$. In the pre-training phase, a feature encoder F_θ is trained on the base class dataset D_b to learn generalizable feature representations while preserving its parameters. In the classifier training phase, few-shot learning tasks (typically called N-way K-shot tasks) are constructed from D_n, where N represents the number of randomly selected classes from C_n, and K indicates the number of samples per class. Each task consists of a support set $S = \{(x_i, y_i)\}_{i=1}^{N\times K}$ for training a new classifier $C(\cdot\,|W_{new})$ to rapidly adapt to novel classes, and a query set $Q = \{(x_i)\}_{i=1}^{N\times q}$ (where q is the number of query samples) for evaluating model performance. In the inference phase, the model is tested on multiple sampled tasks to calculate its average classification accuracy on query sets. The core objective of few-shot learning is to transfer knowledge from base classes to generalize to data-scarce novel classes. By learning general feature representations during pretraining and rapidly adapting a new classifier during classifier training, the model can effectively classify unseen categories when provided with only a few samples.

3.2. *Overview of the framework*

As illustrated in Figure 2, the proposed framework consists of three key phases: pre-training, classifier training, and inference. During the pre-training phase, a dual-task framework combining supervised and self-supervised learning enhances the feature encoder's discriminative and generalization capabilities, enabling it to capture both category labels and data structural information while providing more robust features for subsequent tasks. In the classifier training phase, the preserved pretrained feature encoder parameters are utilized to train new classifiers for few-shot tasks, where features are transformed into class-conditional Gaussian distributions to dynamically calibrate novel class distributions while generating additional sample features to enrich classifier inputs. The inference phase freezes both the feature encoder and classifier for efficient query sample classification,

with a comprehensive assessment of the model's generalization performance in data-scarce scenarios achieved through multi-task testing and average classification result computation.

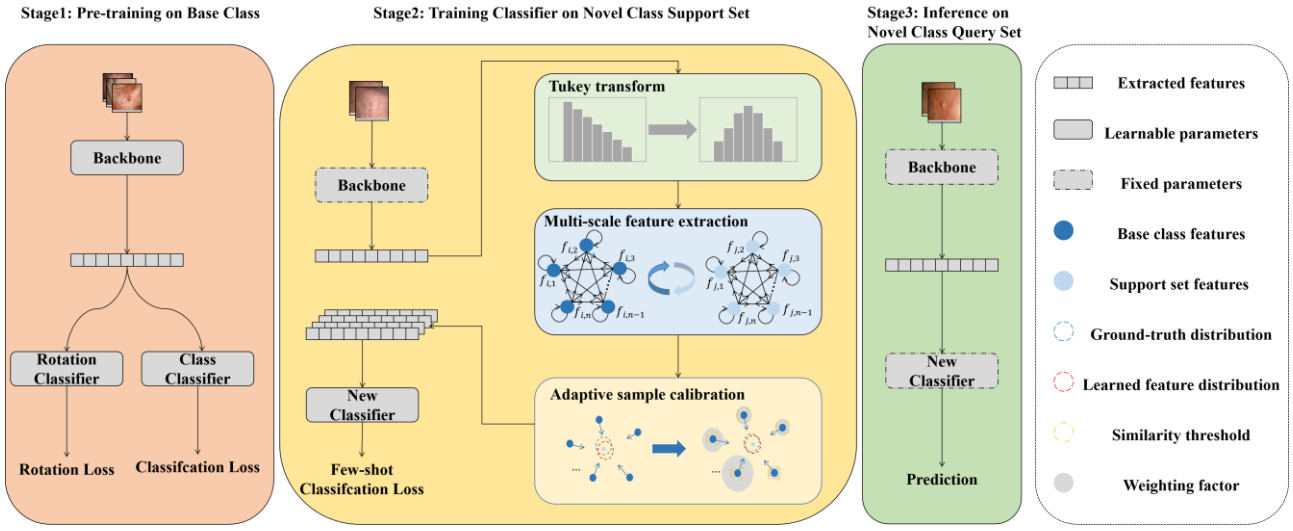

Figure 2. Overall framework.

3.3. Pre-training phase

During the pre-training phase, the feature encoder F_θ is trained on the base class dataset D_b while preserving its parameters to learn generalizable feature representations. Concurrently, statistical information from base classes is retained to calibrate the distribution of novel classes, thereby optimizing classifier inputs during the training phase. Recent studies [31-32] demonstrate that incorporating self-supervised learning during pretraining represents an efficient approach that significantly enhances model generalization without requiring additional annotation. As an auxiliary task, self-supervised learning facilitates the acquisition of more robust feature representations, consequently improving model performance in image understanding tasks. Building on this advantage, we integrate a self-supervised learning mechanism with supervised learning to further enhance overall model performance.

3.3.1. Supervised learning branch

For an input image x_i in the supervised learning branch, its predicted class label $\hat{y}_i$ is computed by:

$$P(y = c|x_i) = Softmax\big(C(F_\theta(x_i)|W_{class})\big) \tag{1}$$

$$\hat{y}_i = arg\max_c P(y = c \mid x_i) \tag{2}$$

where $F_\theta(\cdot)$ denotes the feature encoder, $C(\cdot\,|W_{class})$ represents the classifier, and c indexes classes in the base dataset D_b.

The supervised classification loss L_{class} is obtained by computing the cross-entropy between predicted probabilities and ground-truth labels:

$$L_{class} = -\frac{1}{N}\sum_{i=1}^{N} log\, P(y = y_i|x_i) \quad (3)$$

3.3.2. Self-supervised learning branch

Rotation prediction [33] serves as an effective self-supervised learning task that enhances a model's visual feature understanding by requiring it to identify the rotation angle applied to input images. This approach is particularly beneficial for skin lesion images due to the inherent rotation-invariance of anatomical structures, where diagnostic characteristics such as asymmetry, border irregularity, and color variegation remain consistent across rotations. By predicting rotations, the backbone network is guided to learn robust, low-level structural features (e.g., edges, textures, and shapes) without relying on manual labels, thereby improving semantic sensitivity and reducing dependence on spurious correlations. This results in a stronger feature extractor that enhances performance in downstream few-shot classification tasks, especially when labeled data is scarce.

In this task, input images are randomly rotated by distinct angles (0°, 90°, 180°, or 270°), forcing the model to recognize both the primary subject and its spatial orientation, thereby improving rotation-invariant feature learning. The implementation involves (1) extracting rotation-aware features through encoder F_θ that capture both semantic content and orientation cues, followed by (2) angle prediction via a 4-way linear classifier $C(\cdot\,|W_{rot})$. The rotation loss L_{rot} (analogous to L_{class}) computes the cross-entropy between predicted and true rotation labels r_i:

$$L_{rot} = -\frac{1}{N}\sum_{i=1}^{N} log\, P(r = r_i|x_i) \quad (4)$$

3.4. Classifier training phase

As illustrated in Figure 2, upon completion of pre-training, the feature encoder's parameters are preserved and subsequently employed to train classifiers for each task sampled from novel classes. To further enhance model performance, this stage introduces an adaptive distribution calibration strategy comprising three key components.

3.4.1. Feature preprocessing based on Tukey's power transformation

Studies [23,24] demonstrated that feature distributions become more calibratable in lower-dimensional spaces compared to original image spaces. We therefore employ the pre-trained feature encoder to transform images into feature vectors. This encoder, jointly trained through supervised learning on base classes and self-supervised pretraining tasks, acquires generalizable representations suitable for image understanding. For analytical simplicity, we assume these features follow Gaussian distributions, where the mean and covariance effectively capture distribution characteristics. However, due to ReLU activation in the encoder, the extracted features may deviate from ideal Gaussian distributions (Figure 3, left).

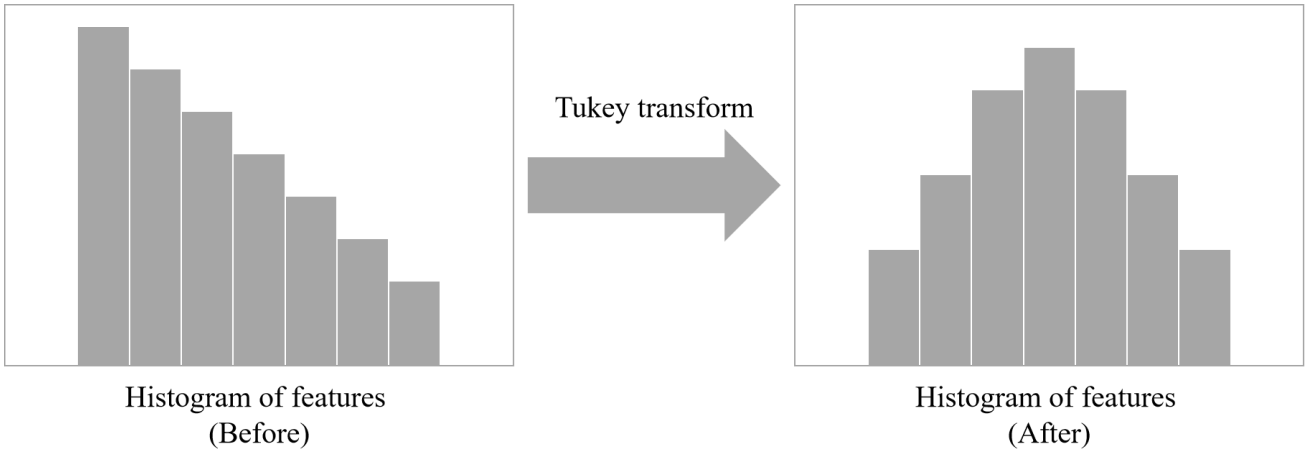

Figure 3. Tukey's power transformation.

To better approximate Gaussian distributions, we apply Tukey's power transformation [34] to both base and novel class features:

$$\tilde{x} = \begin{cases} x^{\alpha}, & \alpha \neq 0 \\ log\ (x), & \alpha = 0 \end{cases} \tag{5}$$

where α controls distribution skewness (reducing α decreases positive skew). The transformed distributions (Figure 3, right) yield calibrated features $\tilde{x}$. For base class j, the mean and covariance are computed as:

$$\mu_j = \frac{1}{n_j} \sum_{i=1}^{n_j} \tilde{x}_i \tag{6}$$

$$\Sigma_j = \frac{\sum_{i=1}^{n_j} (\tilde{x}_i - \mu_j)(\tilde{x}_i - \mu_j)^T}{n_j - 1} \tag{7}$$

where $\tilde{x}_i$ denotes the transformed feature of the i-th sample in class j (n_j = sample count). For N-way K-shot tasks, $\tilde{x}_s$ and $\tilde{x}_q$ represent transformed support and query features, respectively, enabling analogous mean/covariance calculation for novel classes.

3.4.2. Multi-scale feature extraction strategy

The process of selecting and transferring statistical information from base to novel classes requires appropriate sample characterization and relationship measurement. Since sample features are multidimensional, characterizing distributions from multiple directional perspectives provides richer distributional and positional information compared to limited dimensional views. Existing distribution calibration methods typically rely on Euclidean distance as a similarity metric, yet this point-to-point measurement may exhibit limitations in capturing complex distribution characteristics. To overcome this limitation, we propose a multi-scale feature extraction strategy based on feature description matrices and composite metrics (Figure 4). Specifically, the feature description matrix is constructed through the cross-product of feature vectors and their transposes, establishing inter-dimensional

relationships that replace the original features for similarity computation. This enables comprehensive representation of features across global, local, and fine-grained scales, ultimately providing a more robust basis for estimating true distributions in the embedding space.

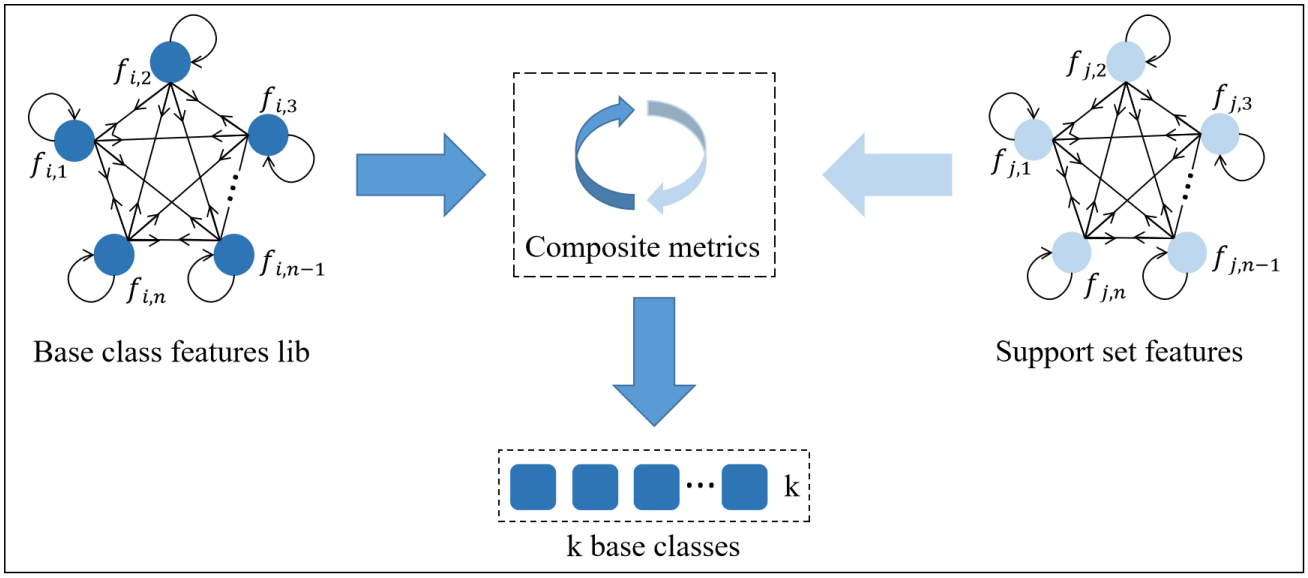

Figure 4. Multi-scale feature extraction strategy.

To more comprehensively characterize feature distributions, we introduce a feature descriptor matrix [35]. This approach captures feature distributions not only along original dimensions but also across other directional perspectives, thereby acquiring more comprehensive distributional information. The feature descriptor matrix is constructed through the outer product of a feature vector and its transpose:

$$M_i = \tilde{x}_i \cdot \tilde{x}_i^T \tag{8}$$

where M_i denotes the feature descriptor matrix, $\tilde{x}_i$ represents the i-th feature vector, and $\tilde{x}_i^T$ is its transpose. Given the larger sample size in base classes, we simplify computation by approximating the base class descriptor matrix as the mean matrix:

$$M_j' = \mu_j \cdot \mu_j^T \tag{9}$$

where M_j' indicates the feature descriptor matrix for the j-th base class, and μ_j is its mean matrix. For support sets with limited samples, we retain the original computation from Equation (8).

For selecting statistically relevant base classes and transferring their information, we introduce a composite metric approach [36] that combines multiple measurement modalities:

$$D_{cm}(M_i, M_j') = \beta \cdot W + \epsilon \cdot \mathbb{F} + \eta \cdot JS \tag{10}$$

Here, β, ϵ, and η are tunable hyperparameters, while W, $\mathbb{F}$, and JS, respectively denote the Wasserstein distance [37], Frobenius norm [38], and Jensen-Shannon (JS) divergence [39]. Specifically:

The Wasserstein distance quantifies distribution dissimilarity:

$$W = \inf_{\gamma \sim \Pi\left(M_i, M_j'\right)} \mathbb{E}_{(x,y)\sim\gamma}[\| x - y \|] \tag{11}$$

The Frobenius norm serves as a generalized Euclidean norm for matrices:

$$\mathbb{F} = -\left\| M_i - M_j' \right\|_F \tag{12}$$

The JS divergence evaluates distribution similarity:

$$JS = \frac{1}{2} KL\left(M_i \,\|\, \frac{M_i + M_j'}{2}\right) + \frac{1}{2} KL\left(M_j' \,\|\, \frac{M_i + M_j'}{2}\right) \tag{13}$$

where KL represents Kullback–Leibler divergence. This composite metric framework enables comprehensive similarity measurement between base and novel classes, significantly improving distribution calibration accuracy through multidimensional relationship characterization.

3.4.3. Adaptive sample calibration strategy and sample augmentation

Following similarity computation via multi-scale feature extraction, conventional approaches typically select the top-k most similar base class samples, which may inadvertently incorporate low-similarity samples and compromise calibration accuracy. To address this, inspired by [40], we propose an adaptive sample calibration strategy consisting of two core components: an adaptive weight matrix and dynamic k-selection (Figure 5), which adaptively selects optimal base class samples for calibrating novel class statistics. The adaptive weight matrix dynamically calculates relevance scores between target novel classes and base classes, while the dynamic k-mechanism determines the optimal number of base samples based on semantic similarity thresholds. This ensures that only the most semantically relevant base class information is utilized for distribution calibration, effectively avoiding noise from irrelevant categories and addressing the challenge of selecting appropriate base class information.

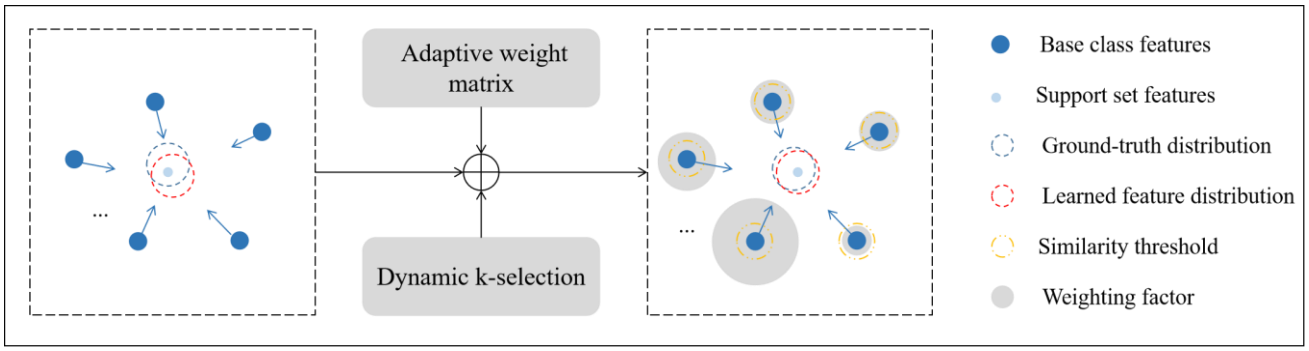

Figure 5. Adaptive sample calibration strategy.

First, based on multi-scale similarity results, the dynamic weight matrix generates normalized weighting coefficients. This matrix adaptively allocates contribution weights to base class samples according to their similarity, avoiding calibration bias caused by fixed weights. Second, to overcome the limitation of fixed k values, we introduce a similarity threshold T for sample screening: only samples with weights exceeding T are retained for calibration. The threshold T is initially set to 0.5

and dynamically adjusted to ensure available samples.

Adaptive weight matrix first constructs a normalized similarity weight matrix that automatically assigns contribution weights to base class samples based on their relevance to each novel class:

$$\lambda_j = \frac{1}{1 + D_{cm}\left(M_i, M_j^{'}\right)} \cdot \mathbb{I}\left(\frac{1}{1 + D_{cm}\left(M_i, M_j^{'}\right)} \geq T\right) \tag{14}$$

where λ_j denotes the weight factor for the j-th base class, and $\mathbb{I}(\cdot)$ is an indicator function activated only when the weight exceeds threshold T (typically initialized at 0.5).

Dynamic k-selection dynamically determines the optimal number of base samples by excluding those with similarity weights below T, ensuring that only highly relevant samples participate in calibration.

The calibrated statistics for novel classes are computed as:

$$\tilde{\mu} = \frac{1}{\sum\lambda_j + 1}\left(\sum\lambda_j\mu_j + \tilde{x}_i\right) \tag{15}$$

$$\tilde{\Sigma} = \frac{1}{\sum\lambda_j + 1}\left(\sum\lambda_j\Sigma_j + \delta\right) \tag{16}$$

where $\tilde{x}_i$ comes from the novel class support set S, μ_j/Σ_j represent base class statistics, and δ compensates for intra-class variation.

The calibrated feature distribution is expressed as:

$$\mathbb{N}_{y_i} = \{\mathcal{N}(\tilde{\mu}_i, \tilde{\Sigma}_i) | i \in (1, \ldots, K)\} \tag{17}$$

where $\tilde{\mu}_i$ and $\tilde{\Sigma}_i$ are calibrated statistics, and K corresponds to the support set size in N-way K-shot settings. The generated feature vectors are:

$$\mathbb{D}_{y_i} = \left\{(x_i, y_i) | x_i \sim \mathcal{N}(\tilde{\mu}_i, \tilde{\Sigma}_i), \forall\mathcal{N}(\tilde{\mu}_i, \tilde{\Sigma}_i) \in \mathbb{N}_{y_i}\right\} \tag{18}$$

The new classifier is trained by minimizing cross-entropy loss:

$$L_{new} = -\frac{1}{\mathbb{D}_{y_i}}\sum_{i=1}^{\mathbb{D}_{y_i}} logP(y = y_i | x_i) \tag{19}$$

3.5. Inference phase

Upon completion of the task-specific classifier training, all parameters are frozen for inference on the query set Q, with the classification probability computed as:

$$P\left(y = n | x_j\right) = Softmax\left(C\left(F_\theta\left(x_j\right) | W_{new}\right)\right), x_j \in Q \tag{20}$$

where $F_\theta\left(x_j\right)$ denotes the query feature representation, and $C(\cdot; W_{new})$ represents the optimized novel-class classifier. The complete workflow of the proposed method is summarized in Algorithm 1.

Algorithm 1 Procedure of SADC

Input: base dataset D_b,novel dataset D_n, Various hyperparameters
Output: Average accuracy and Average AUC for n_{tasks} few-shot tasks
/ Pre-training phase:*
Randomly initialize model weights θ
For each epoch n from 1 to n_{epoch}
    For each batch i from 1 to batchsize
        Randomly sample(x_i, y_i) from D_b
        Obtain rotation set R_i
        Compute total Loss: L_{total}=L_{class}+L_{rot}
    End for
    Update model weights θ: $\theta \leftarrow \theta - \sigma \cdot \nabla_\theta L_{total}$
End for
/ Classifier-training phase:*
Frozen backbone's parameter θ
Extract base classes and novel classes features: x_b=$F_\theta(D_b)$, x_n=$F_\theta(D_n)$
Apply Tukey transform to x_b Obtain the statistics of base classes:$\tilde{x}_b$←Tukey transform (x_b)
For each task i from 1 to n_{task}
    Randomly sample task(x_{S_i}, x_{Q_i}) from x_n
    Apply Tukey transform to task(x_{S_i}, x_{Q_i}) Obtain the statistics of task:
    $\tilde{x}_{S_i}$←Tukey transform (x_{S_i}), $\tilde{x}_{Q_i}$←Tukey transform (x_{Q_i}),
    $\mathbb{N}_{y_i}$=Adaptive Distribution Calibration$(\tilde{x}_{S_i}, \tilde{x}_b)$
    Generate $N_{generate}$ samples for each class n, denoted as $\mathbb{D}_{y_i}$.
    Trained a new classifier on $\mathbb{D}_{y_i} = \{(x_i, y_i) | x_i \sim \mathcal{N}(\tilde{\mu}_i, \tilde{\Sigma}_i), \forall \mathcal{N}(\tilde{\mu}_i, \tilde{\Sigma}_i) \in \mathbb{N}_{y_i}\}$
/ Inference phase:*
    Frozen the classifier's parameters
    Calculate accuracy and AUC on query set $\tilde{x}_{Q_i}$
End for
Compute the average accuracy and average AUC for n_{task} few-shot tasks

4. Experiments

4.1. Datasets

To validate the effectiveness of the proposed method, comprehensive experiments were conducted on three publicly available dermatology datasets. The datasets were carefully selected to represent diverse clinical scenarios and diagnostic challenges in skin disease recognition.

ISIC2018: The ISIC2018 dataset [41] comprises 10,015 dermatological images across 7 disease categories, with each image originally sized at 600×450 pixels. Following established protocols [10,30], we partitioned the dataset into four base classes (containing more abundant samples) and three novel classes (with fewer samples) as illustrated in Figure 1(a). This partitioning scheme enables rigorous evaluation of few-shot learning performance on clinically relevant but data-scarce conditions.

Derm7pt: Containing 2005 high-resolution images (768×512 pixels) spanning 20 diagnostic categories [42], the Derm7pt dataset was processed by excluding the "Miscellaneous" (as it represents

random skin conditions not belonging to any defined disease class) and "Melanoma" (due to having only a single sample, which prevents a meaningful train-test split) categories per prior research [10,30]. From the remaining 18 clinically distinct conditions, we designated 13 well-represented classes as base categories and 5 sparsely sampled classes as novel categories [Figure 1(b)], creating a challenging testbed for evaluating diagnostic generalization.

SD-198: As one of the most comprehensive dermatology resources, the SD-198 dataset [43] includes 6584 high-definition images (1640×1130 pixels) across 198 fine-grained skin disease categories. Maintaining consistency with benchmark studies [10,30], we established 128 base classes and 70 novel classes [Figure 1(c)], providing an extensive evaluation framework for assessing model performance across a wide spectrum of dermatological conditions with varying prevalence.

4.2. Implementation details

The experiments were conducted in the following environment: an NVIDIA GeForce RTX 4090 GPU (24GB VRAM), PyTorch 2.1.0 deep learning framework, Python 3.10 programming language, and CUDA 12.1. To maintain input consistency, all images were uniformly resized to 80×80 pixels. The WRN28 network [44] with integrated rotation prediction was employed as the feature encoder during pretraining. The Adam optimizer [45] was utilized for parameter updates with an initial learning rate σ=0.001 and weight decay of 1e-4 to prevent overfitting. The model underwent 100 training epochs with a fixed batch size of 32, with optimal weights preserved for transfer to subsequent classifier training phases. For classifier training, a novel logistic regression classifier was trained and evaluated on new categories. All models were evaluated under the same category-wise data splitting scheme to ensure a fair comparison, following the standard evaluation protocol established in few-shot learning literature. To evaluate computational efficiency, we measured the runtime of our complete method under standard 2-way 1-shot settings, which processes individual tasks in 0.931 seconds on an NVIDIA GeForce RTX 4090 GPU. The adaptive feature distribution calibration employed the following hyperparameters: Tukey power transformation exponent α=0.5, Wasserstein distance weight β=1, Frobenius norm weight ϵ=1, JS divergence weight η=0.5, similarity threshold T=0.4, and intra-class variation compensation δ=0.01. Note that these hyperparameters (α=0.5, β=1, ϵ=1, η=0.5, and T=0.4) were selected through grid search, as detailed in Section 4.5. Following established methodologies [10,30], 600 tasks were randomly sampled from the novel class dataset for testing to ensure experimental fairness and comparability. Model performance was comprehensively evaluated through mean accuracy and average AUC calculations.

4.3. Results

4.3.1. Results on ISIC2018

Table 1 presents the comparative results of 2-way classification tasks on the ISIC2018 dataset. The proposed SADC method demonstrates superior performance over current state-of-the-art approaches, achieving significant improvements in both AUC and accuracy metrics. Specifically, for 2-way 1-shot, 3-shot, and 5-shot tasks, SADC yields AUC improvements of 2.6%, 1.64%, and 1.68% respectively, while corresponding accuracy gains reach 1.67%, 2.75%, and 1.16%.

Table 2 further illustrates the method's robustness in more challenging 3-way classification

scenarios. Following established evaluation protocols, performance was assessed on 400 randomly sampled tasks from novel classes, with classification accuracy as the primary metric. Compared to existing state-of-the-art methods, SADC achieves accuracy improvements of 1.19%, 0.66%, and 0.5% for 3-shot, 5-shot, and 10-shot tasks, respectively, consistently demonstrating its competitive advantage across varying difficulty levels.

Table 1. Performance comparison of 2-way classification tasks on ISIC2018.

Method	2-way 1-shot		2-way 3-shot		2-way 5-shot	
	Avg. AUC	Avg. Acc	Avg. AUC	Avg. Acc	Avg. AUC	Avg. Acc
Reptile	60.3	58.0	73.1	73.4	79.6	76.2
ProtoNet	61.6	59.3	70.2	67.9	75.4	73.0
Meta-Derm	68.1	64.3	81.2	76.7	86.8	82.1
Meta-Rep	72.6	65.9	77.4	76.5	80.1	79.6
MAML	61.72	60.16	73.38	74.56	78.29	79.16
ST-Meta	65.27	65.78	76.32	76.38	80.59	81.38
SS-DCN	75.77	68.45	85.94	79.22	88.88	82.63
Ours	**78.37±0.75**	**70.12±0.66**	**87.58±0.67**	**81.97±0.33**	**90.56±0.72**	**83.79±0.43**

Table 2. Performance comparison of 3-way classification tasks on ISIC2018.

Method	3-way 3-shot	3-way 5-shot	3-way 10-shot
	Avg. Acc	Avg. Acc	Avg. Acc
MetaMed	58.50	61.25	71.00
Transfer	55.67	59.67	65.92
PT-MAP	53.17	55.61	59.57
Baseline	56.80	59.20	65.22
PFEMed	66.94	69.78	73.81
ST-Meta	59.79	64.59	–
SS-DCN	66.34	70.69	74.79
Ours	**68.13±1.17**	**71.35±1.09**	**75.29±1.01**

4.3.2. Results on Derm7pt

Table 3 presents the comparative results of 2-way classification tasks on the Derm7pt dataset. The proposed SADC method achieves state-of-the-art performance across all experimental settings. Specifically, for 2-way 1-shot, 3-shot, and 5-shot tasks, SADC demonstrates AUC improvements of 0.36%, 0.51%, and 0.16%, respectively, while achieving corresponding accuracy gains of 0.61%, 1.14%, and 1.18%.

The limited scale of the Derm7pt dataset makes models particularly susceptible to overfitting. SADC effectively addresses this challenge through its innovative integration of self-supervised learning as a regularizer. Notably, compared to SCAN, SADC delivers substantial accuracy

improvements of 6.51% and 13.2% for 2-way 1-shot and 5-shot tasks, respectively. While SCAN enhances performance by learning sub-cluster structures within datasets, its effectiveness remains highly dependent on dataset characteristics, showing limited generalization on Derm7pt. In contrast, SADC's architecture demonstrates superior adaptability to diverse dataset properties, exhibiting more robust generalization capabilities across different diagnostic scenarios.

Table 3. Performance comparison of 2-way classification tasks on Derm7pt.

Method	2-way 1-shot		2-way 3-shot		2-way 5-shot	
	Avg. AUC	Avg. Acc	Avg. AUC	Avg. Acc	Avg. AUC	Avg. Acc
Reptile	59.7	60.2	64.1	65.7	71.4	70.5
ProtoNet	60.6	62.5	65.8	63.9	68.2	66.7
Meta-Derm	62.1	61.8	68.7	69.9	77.2	76.9
Meta-Rep	72.9	64.0	78.6	74.3	83.2	78.1
PCN	–	59.98	–	–	–	70.62
SCAN	–	61.42	–	–	–	72.58
SS-DCN	74.89	67.32	86.83	78.13	92.87	84.60
Ours	**75.25±0.42**	**67.93±0.3**	**87.34±0.36**	**79.27±0.24**	**93.03±0.31**	**85.78±0.48**

4.3.3. Results on SD-198

Table 4 presents the comparative results of 2-way classification tasks on the SD-198 dermatology dataset. The experimental results demonstrate that SADC outperforms previous state-of-the-art methods in both AUC and accuracy metrics. Specifically, for 2-way 1-shot, 3-shot, and 5-shot tasks, SADC achieves AUC improvements of 1.16%, 1.02%, and 1.25%, respectively, along with corresponding accuracy gains of 1.71%, 0.83%, and 0.92%.

Table 4. Performance comparison of 2-way classification tasks on SD-198.

Method	2-way 1-shot		2-way 3-shot		2-way 5-shot	
	Avg. AUC	Avg. Acc	Avg. AUC	Avg. Acc	Avg. AUC	Avg. Acc
Reptile	64.1	63.0	77.4	72.9	84.6	80.4
ProtoNet	59.4	59.8	70.6	66.6	80.7	78.3
Meta-Derm	68.6	65.3	79.1	75.8	89.5	83.7
IPNet	–	–	83.00	78.41	87.00	84.20
Baseline	–	73.98	–	–	–	88.67
SCAN	–	77.12	–	–	–	90.22
SS-DCN	85.63	78.52	93.69	87.53	95.81	90.43
Ours	**86.79±0.79**	**80.23±0.47**	**94.71±0.42**	**88.36±0.38**	**97.06±0.38**	**91.35±0.26**

Overall, SADC demonstrates significantly superior accuracy compared to SS-DCN. As discussed earlier, SS-DCN focuses solely on determining what to transfer (the content of distribution information)

while neglecting how much to transfer (the adaptive weighting of base class contributions). In contrast, SADC's adaptive distribution calibration strategy generates feature distributions that better approximate the true underlying distributions, thereby achieving enhanced classification performance.

4.4. Ablation study

This section conducts systematic ablation experiments on the ISIC2018 dataset, performing in-depth investigations from multiple perspectives, including module effectiveness validation, feature encoder verification, and classifier performance analysis. Through these comprehensive experiments, we thoroughly evaluate the impact of each component on model performance and deeply analyze their respective contributions and operational mechanisms.

4.4.1. Module effectiveness verification

Table 5 presents comprehensive ablation results evaluating the performance contributions of key components in the SADC method. The study specifically examines the impacts of: (1) Tukey's power transformation, (2) multi-scale feature extraction strategy, and (3) adaptive sample calibration strategy on the ISIC2018 dataset.

The baseline model (without any SADC components) achieves fundamental performance metrics. Comparative analysis reveals that for 2-way 1-shot tasks:

a). Tukey transformation yields AUC/accuracy improvements of 4.4%/3.93%.

b). Multi-scale feature extraction contributes with 0.96%/0.95% gains.

c). Adaptive sample calibration provides 1.89%/1.94% enhancements.

Notably, Tukey's transformation demonstrates particularly significant accuracy improvements, confirming its effectiveness in distribution normalization. The experimental results further establish that combined module integration produces substantially better outcomes than individual component usage. The complete SADC framework (integrating all three modules) achieves remarkable performance boosts of 10.84% AUC and 6.44% accuracy for 2-way 1-shot tasks.

Table 5. Module effectiveness validation.

Tukey transform	Multi-scale feature extraction	Adaptive sample calibration	2-way 1-shot	
			Avg. AUC	Avg. Acc
✗	✗	✗	67.53±0.55	63.68±0.3
✓	✗	✗	71.93±0.67	67.61±0.48
✗	✓	✗	68.49±0.46	64.63±0.35
✗	✗	✓	69.42±0.41	65.62±0.61
✓	✓	✗	74.97±0.58	68.55±0.39
✗	✓	✓	71.85±0.49	67.44±0.37
✓	✗	✓	75.34±0.59	68.84±0.4
✓	✓	✓	**78.37±0.34**	**70.12±0.29**

The findings quantitatively validate each module's effectiveness while demonstrating critical synergistic effects between components, substantiating the proposed method's architectural superiority. The ablation study provides conclusive evidence that SADC's performance gains stem from strategic

feature space transformation, comprehensive distribution characterization, and adaptive calibration mechanisms, with this systematic verification confirming that our methodological innovations collectively address key challenges in few-shot dermatological image analysis.

4.4.2. Feature encoder effectiveness verification

The SADC method demonstrates strong compatibility with various feature encoders. To investigate the impact of encoder selection, Table 6 presents 2-way 1-shot classification results on the ISIC2018 dataset using five distinct architectures: a 4-layer CNN (conv4), 6-layer CNN (conv6), ResNet18, WRN28, and WRN28 enhanced with rotation prediction pretraining. Experimental results reveal two key findings: (1) more sophisticated encoders significantly boost downstream classification performance, and (2) self-supervised learning through rotation prediction enables encoders to extract more discriminative and generalizable features by leveraging inherent structures in unlabeled data. Specifically, the rotation-pretrained WRN28 achieves remarkable improvements over other encoders, with AUC gains of 9.84%, 7.69%, 6.64%, and 3.76% respectively, alongside corresponding accuracy improvements of 5.43%, 4.35%, 3.56%, and 1.89% in comparative evaluations.

Table 6. Feature encoder effectiveness validation.

Backbones	2-way 1-shot	
	Avg. AUC	Avg. Acc
Conv4	68.53±0.34	64.69±0.27
Conv6	70.68±0.38	65.77±0.31
Resnet18	71.73±0.34	66.56±0.24
WRN28	74.61±0.44	68.73±0.37
WRN28+Rotation Loss	**78.37±0.47**	**70.12±0.38**

4.4.3. Classifier effectiveness verification

The SADC method demonstrates classifier-agnostic characteristics due to its feature-level operations, while still exhibiting significant performance variations across different classifier choices. Our comprehensive ablation study on classifier selection reveals that the logistic regression classifier employed in SADC outperforms alternative approaches, including support vector machines (with both RBF and Linear kernels) and Naïve Bayes classifiers (Table 7). Quantitative results from 2-way 1-shot tasks show that the logistic regression classifier achieves superior performance with AUC improvements of 6.16%, 2.78%, and 7.59% compared with RBF-SVM, linear-SVM, and Naïve Bayes classifiers, respectively, accompanied by corresponding accuracy gains of 3.16%, 0.9%, and 3.93%. These findings not only validate the effectiveness of logistic regression within the SADC framework but also highlight its crucial role in enhancing overall model performance through optimal feature space utilization.

Table 7. Classifier effectiveness validation.

Classifiers	2-way 1-shot	
	Avg. AUC	Avg. Acc
SVM (kernel = rbf)	72.21±0.55	66.96±0.36
SVM (kernel = linear)	75.59±0.42	69.22±0.4
Naive Bayes	70.78±0.36	66.19±0.29
Logistic regression	**78.37±0.34**	**70.12±0.3**

4.5. Hyperparameter tuning

To investigate the impact of hyperparameter settings on model performance, we conducted systematic hyperparameter tuning experiments on the ISIC2018 dataset to identify the optimal configuration that maximizes classification accuracy. Figure 6 presents the tuning results for 2-way 1-shot tasks, evaluating five critical hyperparameters: the exponent α in Tukey's power transformation, weighting coefficients β (Wasserstein distance), ϵ (Frobenius norm), η (JS divergence), and similarity threshold T.

The experimental results demonstrate that SADC achieves peak accuracy in 2-way 1-shot tasks when configured with α=0.5, β=1, ϵ=1, η=0.5, and T=0.4. Notably, Figure 6 reveals that among the three distance metrics, JS divergence (η) contributes relatively less to performance improvement, as evidenced by its lowest optimal weight. Furthermore, this hyperparameter configuration enables SADC to surpass state-of-the-art results across all three benchmark datasets, confirming both the stability of the parameter optimization process and the robustness of the proposed method under different evaluation scenarios.

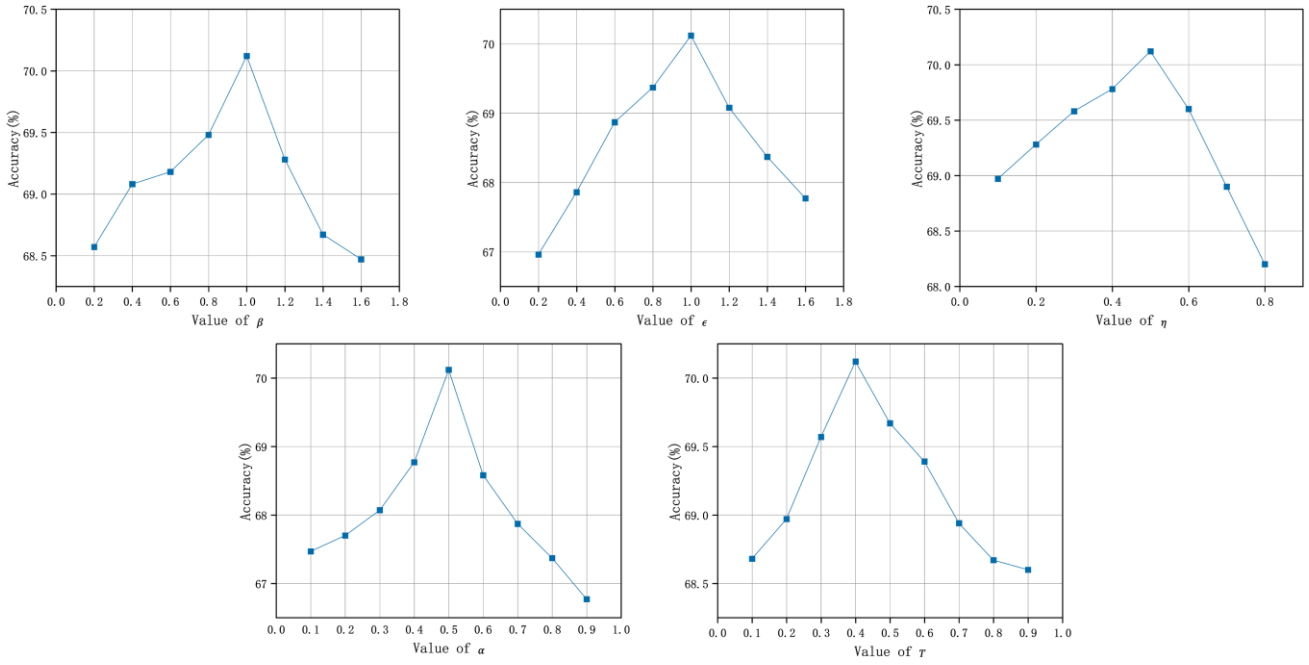

Figure 6. Hyperparameter tuning.

4.6. Generate feature visualization

Figure 7 presents t-SNE [46] visualization results for a randomly selected 2-way 1-shot task from the ISIC2018 dataset, comparing ground-truth distributions with feature samples generated by different methods. The analysis focuses on four key scenarios: (a) the original support set distribution (containing only 2 samples), (b) features generated by our adaptive distribution calibration (ADC) approach, (c) features produced by conventional distribution calibration (DC), and (d) the actual ground-truth feature distribution.

The visualization clearly demonstrates that while DC-generated features exhibit noticeable deviation from the true distribution [Figure 7(c)], our ADC method achieves significantly better alignment with the ground-truth characteristics (Figure 7(b)). This compelling visual evidence validates ADC's ability to reconstruct more authentic feature distributions, which directly explains the method's superior performance in few-shot classification tasks while simultaneously revealing the fundamental limitations of conventional DC approaches that fail to maintain comparable distributional fidelity.

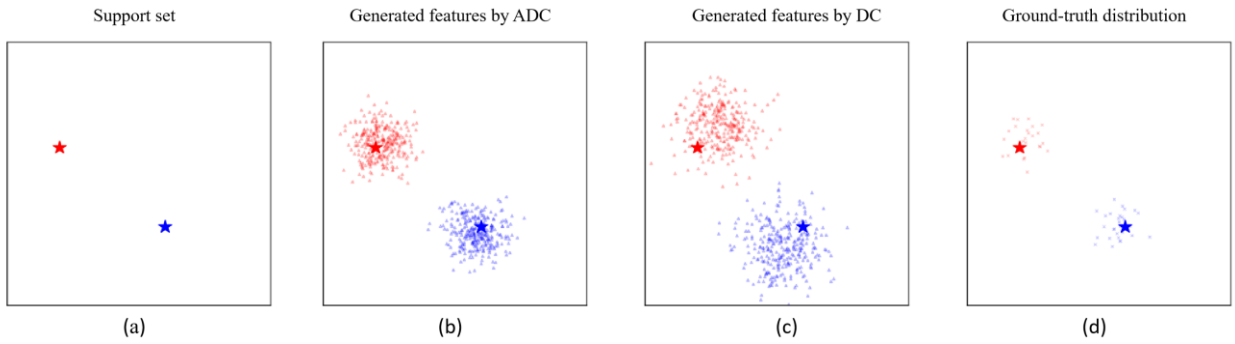

Figure 7. t-SNE visualization of 2-way 1-shot tasks on ISIC-2018.

5. Conclusions

This paper proposes SADC, an adaptive distribution calibration-based few-shot learning method for dermatological classification, which significantly improves diagnostic accuracy for rare skin conditions in data-scarce scenarios. The framework incorporates two key innovations: (1) a multi-scale feature extraction strategy that enhances base-to-novel class similarity measurement through multidimensional feature extraction, and (2) an adaptive sample calibration mechanism that intelligently selects relevant base samples and optimizes feature distribution generation to better approximate the true data distribution. Comprehensive experiments demonstrate that SADC achieves substantial performance improvements over state-of-the-art approaches across multiple evaluation metrics.

This study acknowledges several limitations: model performance requires further improvement for disease categories with fundamentally novel pathological characteristics, and validation currently relies primarily on public datasets rather than real-world clinical data. This study adopted the category-wise data splitting protocol consistent with prior works. However, since patient identifiers were unavailable in the datasets used, patient-wise data splitting could not be performed. Consequently, images from the same patient might exist in both training and testing sets, potentially introducing

patient-specific biases and overestimating the model's generalization performance. Therefore, the results should be interpreted with caution. Future research should seek datasets with patient-level annotations to enable more rigorous validation.

Our future work will implement a two-stage validation pathway: first, conducting in-depth technical validation through generalized meta-learning strategies, followed by clinical validation in collaboration with partner hospitals. Dermatologists will perform comparative assessments to quantify the model's actual effectiveness in improving diagnostic accuracy and efficiency for rare diseases. Furthermore, we will advance integration with clinical decision support systems by developing HIS-compliant API interfaces, conducting rigorous FDA/CE certification testing, and creating explainable AI interfaces, ultimately culminating in clinical pilots at local tertiary hospitals.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This study was supported by the National Natural Science Foundation of Hunan Province, China (Grant no. 2025JJ70071, 2025JJ70028, 2025JJ81178, 2024JJ9550) and the Scientific Research Fund of the Hunan Provincial Education Department, China (Grant no. 24A0401).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. F. Grignaffini, F. Barbuto, L. Piazzo, M. Troiano, P. Simeoni, F. Mangini, et al., Machine learning approaches for skin cancer classification from dermoscopic images: A systematic review, *Algorithms*, **15** (2022), 438. https://doi.org/10.3390/a15110438
2. K. M. Hosny, W. Said, M. Elmezain, M. A. Kassem, Explainable deep inherent learning for multi-classes skin lesion classification, *Appl. Soft Comput.*, **159** (2024), 111624. https://doi.org/10.1016/j.asoc.2024.111624
3. D. Meedeniya, S. D. Silva, L. Gamage, U. Isuranga, Skin cancer identification utilizing deep learning: A survey, *IET Image Process.* **18** (2024), 3731–3749. https://doi.org/10.1049/ipr2.13219
4. H. Naseri, A. A. Safaei, Diagnosis and prognosis of melanoma from dermoscopy images using machine learning and deep learning: a systematic literature review, *BMC Cancer*, **25** (2025), 75. https://doi.org/10.1186/s12885-024-13423-y
5. Y. Song, T. Wang, P. Cai, S. K. Mondal, J. P. Sahoo, A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities, *ACM Comput. Surveys*, **55** (2023), 1–40. https://doi.org/10.1145/3582688
6. W. Zeng, Z. Y. Xiao, Few-shot learning based on deep learning: A survey, *Math. Biosci. Eng.,* **21** (2024), 679–711. https://doi.org/10.3934/mbe.2024029

7. A. Parnami, M. Lee, Learning from few examples: A summary of approaches to few-shot learning. Preprint, (2022), arXiv:2203.04291.
8. C. Finn, P. Abbeel, S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, In: *Proceedings of the 34th International Conference on Machine Learning*, (2017), 1126–1135.
9. J. Snell, K. Swersky, R. Zemel, Prototypical networks for few-shot learning, *Adv. Neural Inform. Process. Syst.*, (2017), 30. http://arxiv.org/abs/1703.05175
10. K. Mahajan, M. Sharma, L. Vig, Meta-dermdiagnosis: Few-shot skin disease identification using meta-learning, In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, (2020), 730–731. https://doi.org/10.1109/CVPRW50498.2020.00373
11. X. Li, L. Yu, Y. Jin, C. W. Fu, L. Xing, P. A. Heng, Difficulty-aware meta-learning for rare disease diagnosis, *Med. Image Comput. Computer Assisted Intervent.*, (2020), 357–366. https://doi.org/10.1007/978-3-030-59710-8_35
12. R. Singh, V. Bharti, V. Purohit, A. Kumar, A. K. Singh, S. K. Singh, MetaMed: Few-shot medical image classification using gradient-based meta-learning, *Pattern Recogn.*, **120** (2021), 108111. https://doi.org/10.1016/j.patcog.2021.108111
13. W. Zhu, H. Liao, W. Li, W. Li, J. Luo, Alleviating the incompatibility between cross entropy loss and episode training for few-shot skin disease classification, *Med. Image Comput. Computer Assisted Intervent.*, (2020), 330–339.
14. R. R. Chowdhury, D. R. Bathula, Influential prototypical networks for few shot learning: A dermatological case study, In: *2022 IEEE 19th International Symposium on Biomedical Imaging*, (2022), 1–4.
15. C. Zhou, M. Sun, L. Chen, A. Cai, J. Fang, Few-shot learning framework based on adaptive subspace for skin disease classification, In: *2022 IEEE International Conference on Bioinformatics and Biomedicine*, (2022), 2231–2237. https://doi.org/10.1109/BIBM55620.2022.9995042
16. J. Xiao, H. Xu, D. Fang, C. Cheng, H. Gao, Boosting and rectifying few-shot learning prototype network for skin lesion classification based on the internet of medical things, *Wireless Networks*, **29** (2023), 1507–1521. https://doi.org/10.1007/s11276-021-02713-z
17. Z. Dai, J. Yi, L. Yan, Q. Xu, L. Hu, Q. Zhang, et al., PFEMed: Few-shot medical image classification using prior guided feature enhancement, *Pattern Recogn.*, **134** (2023), 109108. https://doi.org/10.1016/j.patcog.2022.109108
18. D. Chen, Y. Chen, Y. Li, F. Mao, Y. He, H. Xue, Self-supervised learning for few-shot image classification, In: *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing*, (2021), 1745–1749. https://doi.org/10.1109/ICASSP39728.2021.9413783
19. C. Medina, A. Devos, M. Grossglauser, Self-supervised prototypical transfer learning for few-shot classification, preprint, arXiv:2006.11325.
20. C. P. Phoo, B. Hariharan, Self-training for few-shot transfer across extreme task differences, preprint, (2020), arXiv:2010.07734.
21. V. Prabhu, A. Kannan, M. Ravuri, M. Chaplain, D. Sontag, X. Amatriain, Few-shot learning for dermatological disease diagnosis, In: *Machine Learning for Healthcare Conference*, (2019), 532–552. https://doi.org/10.1016/B978-0-32-399851-2.00022-3

22. S. Li, X. Li, X. Xu, K. T. Cheng, Dynamic subcluster-aware network for few-shot skin disease classification, *IEEE Transact. Neural Networks Learn. Syst.*, (2023). https://doi.org/10.1109/TNNLS.2023.3336765
23. S. Yang, L. Liu, M. Xu, Free lunch for few-shot learning: Distribution calibration, preprint, (2021), arXiv:2101.06395.
24. Y. Xian, T. Lorenz, B. Schiele, Z. Akata, Feature generating networks for zero-shot learning, In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2018), 5542–5551. https://doi.org/10.1109/CVPR.2018.00581
25. A. Nichol, J. Schulman, Reptile: A scalable metalearning algorithm, preprint, (2018), arXiv:1803.02999.
26. O. Vinyals, C. Blundell, T. Lillicrap, D. Wierstra, Matching networks for one shot learning, *Adv. Neural Inform. Process. Syst.*, **29** (2016). http://arxiv.org/abs/1606.04080
27. F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. Torr, T. M. Hospedales, Learning to compare: Relation network for few-shot learning, In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2018), 1199–1208. https://doi.org/10.1109/CVPR.2018.00131
28. W. Y. Chen, Y. C. Liu, Z. Kira, Y. C. F. Wang, J. B. Huang, A closer look at few-shot classification, preprint, (2019), arXiv:1904.04232.
29. P. Mangla, N. Kumari, A. Sinha, M. Singh, B. Krishnamurthy, V. N. Balasubramanian, Charting the right manifold: Manifold mixup for few-shot learning, In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, (2020), 2218–2227. https://doi.org/10.1109/WACV45572.2020.9093338
30. W. Fu, J. Chen, L. Zhou, Boosting few-shot rare skin disease classification via self-supervision and distribution calibration, *Biomed. Eng. Letters*, **14** (2024), 877–889. https://doi.org/10.1007/s13534-024-00383-2
31. J. Gui, T. Chen, J. Zhang, Q. Cao, Z. Sun, H. Luo, D. Tao, A survey on self-supervised learning: Algorithms, applications, and future trends, *IEEE Transact. Pattern Anal. Mach. Intell.*, (2024). https://doi.org/10.1109/TPAMI.2024.3415112
32. E. Jang, C. Devin, V. Vanhoucke, S. Levine, Grasp2vec: Learning object representations from self-supervised grasping, preprint, (2018), arXiv:1811.06964.
33. S. Gidaris, P. Singh, N. Komodakis, Unsupervised representation learning by predicting image rotations, preprint, (2018), arXiv:1803.07728.
34. J. W. Tukey, Exploratory Data Analysis, *Reading, MA: Addison-Wesley*, **2** (1977), 131–160.
35. X. Wei, W. Du, H. Wan, W. Min, Feature distribution fitting with direction-driven weighting for few-shot images classification, In: *Proceedings of the AAAI Conference on Artificial Intelligence*, **37** (2023), 10315–10323. https://doi.org/10.1609/aaai.v37i9.26228
36. L. Chen, Y. Gu, Y. Guo, F. Dong, D. Jiang, Y. Chen, DDC: Dynamic distribution calibration for few-shot learning under multi-scale representation, *Knowledge-Based Syst.*, (2025), 113030. https://doi.org/10.1016/j.knosys.2025.113030
37. M. Arjovsky, S. Chintala, L. Bottou, Wasserstein generative adversarial networks, In: *International Conference on Machine Learning*, (2017), 214–223.
38. A. Brauer, J. E. Shockley, On a problem of Frobenius, *J. Reine Angew. Math.*, **211** (1962), 215–220. https://doi.org/10.1515/crll.1962.211.215
39. M. L. Menéndez, J. A. Pardo, L. Pardo, M. D. C. Pardo, The Jensen-Shannon divergence, *J. Franklin Institute*, **334** (1997), 307–318. https://doi.org/10.1016/S0016-0032(96)00063-4

40. X. Liu, K. Zhou, P. Yang, L. Jing, J. Yu, Adaptive distribution calibration for few-shot learning via optimal transport, *Inform. Sci.*, **611** (2022), 1–17. https://doi.org/10.1016/j.ins.2022.07.189
41. N. C. F. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, et al., Skin lesion analysis toward melanoma detection: A challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), hosted by the International Skin Imaging Collaboration (ISIC), In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, (2018), 168–172.
42. J. Kawahara, S. Daneshvar, G. Argenziano, G. Hamarneh, Seven-point checklist and skin lesion classification using multitask multimodal neural nets, *IEEE J. Biomed. Health Inform.*, **23** (2018), 538–546. https://doi.org/10.1109/JBHI.2018.2824327
43. X. Sun, J. Yang, M. Sun, K. Wang, A benchmark for automatic visual classification of clinical skin disease images, In: *Computer Vision–ECCV 2016: 14th European Conference, Proceedings, Part VI 14*, (2016), 206–222. https://doi.org/10.1007/978-3-319-46466-4_13
44. S. Zagoruyko, N. Komodakis, Wide residual networks, preprint, (2016), arXiv:1605.07146.
45. D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, preprint, (2014), arXiv:1412.6980.
46. L. Van der Maaten, G. Hinton, Visualizing data using t-SNE, *J. Mach. Learn. Res.*, **9** (2008).
47. L. Gamage, U. Isuranga, S. De Silva, D. Meedeniya, Melanoma skin cancer classification with explainability, In: *Proceedings of the 2023 3rd International Conference on Advanced Research in Computing (ICARC)*, (2023), 30–35. https://doi.org/10.1109/ICARC57651.2023.10145622
48. L. Gamage, U. Isuranga, D. Meedeniya, S. De Silva, Melanoma skin cancer identification with explainability utilizing mask guided technique, *Electronics*, **13** (2024), 680. https://doi.org/10.3390/electronics13040680

AIMS Press