Open Access Full Text Article

ORIGINAL RESEARCH

Few-Shot Classification with Multiscale Feature Fusion for Clinical Skin Disease Diagnosis

Tianle Chen, Qi Liu, Jie Yang

College of Biomedical Engineering, Sichuan University, Chengdu, 610065, People's Republic of China

Correspondence: Qi Liu, College of Biomedical Engineering, Sichuan University, Chengdu, 610065, People's Republic of China, Email liuqi@scu.edu.cn

Introduction: Skin disease is one of the most common diseases and can affect people of all ages and races. However, the diagnosis of skin diseases via observation is a highly challenging task for both doctors and patients, and would benefit from the use of an intelligent system. Building a large benchmark with professional dermatologists is resource-intensive, and we believe that few-shot learning (FSL) methods would be helpful in solving the problem of annotated data scarcity. In this paper, we propose CDD-Net (Context Feature Fusion and Dual Attention Dermatology Net), a plug-in module for FSL clinical skin disease classification.

Methods: Current FSL methods used in skin disease classification are limited to nonuniversal approaches and few disease classes. Our CDD-Net has a flexible structure, including a context feature–fusion module and dual-attention module to extract discriminating texture feature and emphasize contributive regions and channels. The context feature–fusion module localizes discriminatory texture details of skin lesions by integrating features from different layers, while the dual-attention module highlights discriminative regions via channel-wise and pixel-wise depictions based on weight vectors and restrains the contributions of irrelevant areas. We also present Derm104, a new clinical skin disease data benchmark that has significant coverage of rare diseases and reliable annotation between primary species and subspecies for better validation of our approach.

Results: Our experiments validated the versatility of CDD-Net for different FSL methods and achieved an improvement in accuracy of up to 9.14 percentage points compared with the vanilla network, which can be considered state of the art. The ablation study also showed that the dual-attention module and context feature–fusion module worked efficiently in CDD-Net.

Keywords: clinical skin image, computer-aided diagnosis, few-shot learning, feature fusion, attention mechanism
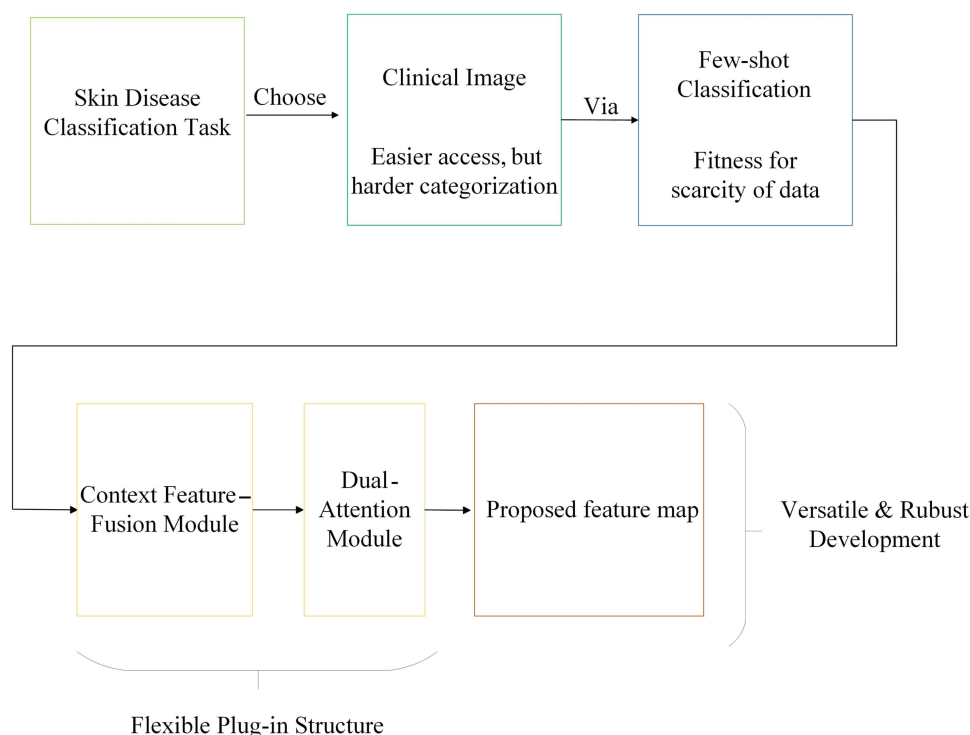
Introduction

Skin disease is one of the most common diseases and can affect people of all ages and races.[1] There are many problems that skin diseases can cause to patients, including itching and bleeding, which can seriously affect their quality of life or even cause them to lose their lives. Skin diseases should be diagnosed early so that they can receive the correct treatment as soon as possible and avoid further progression. Delayed diagnosis can be attributed to the limited medical knowledge of patients and disparity in medical resources. We can solve some of these problems with the help of computer-aided diagnosis to a certain extent.

Dermoscopic images are the primary focus of early investigations using computer-aided diagnosis for skin diseases.[2,3] This is due to the fact that they focus more on lesions than clinical images with uniform illumination and less noise. However, dermoscopic-based diagnosis of skin diseases has some restrictions, such as high costs and low convenience. In recent years, some researchers have begun to pay more attention to clinical images.[4,5] Differences between clinical images and other images are demonstrated in Figure 1.

In recent years, the fields of feature learning and object recognition have experienced tremendous growth in the use of convolution neural networks (CNNs). According to numerous studies from ImageNet's large-scale visual recognition challenge, the most sophisticated CNN has outperformed humans on object-classification tasks.[6–8] Due to its exceptional performance over traditional methods, deep CNN-based learning is also frequently utilized in skin disease classification,[9,10] lesion localization, and segmentation tasks,[11–14] A majority of these tasks showed a high standard of accuracy, making automatic skin disease screening achievable.

Graphical Abstract

```
┌─────────────────┐        ┌─────────────────┐       ┌─────────────────┐
│                 │        │  Clinical Image │       │   Few-shot      │
│  Skin Disease   │ Choose │                 │  Via  │ Classification  │
│ Classification  │───────▶│  Easier access, │──────▶│                 │
│     Task        │        │  but harder     │       │   Fitness for   │
│                 │        │  categorization │       │  scarcity of    │
└─────────────────┘        └─────────────────┘       │     data        │
                                                     └─────────────────┘
```

```
┌──────────────┐     ┌──────────┐     ┌──────────────────┐
│   Context    │     │ Dual-    │     │                  │      Versatile & Rubust
│   Feature-   │────▶│Attention │────▶│ Proposed feature │         Development
│ Fusion Module│     │ Module   │     │      map         │
│              │     │          │     │                  │
└──────────────┘     └──────────┘     └──────────────────┘

         Flexible Plug-in Structure
```

Owing to the great amount of resources required for building a large benchmark with professional dermatologists, we believe that few-shot learning (FSL) classification would be helpful in solving the problem of scarcity of annotated data.[15] It aims to accommodate novel classes unseen during training using only a few examples during testing. This is
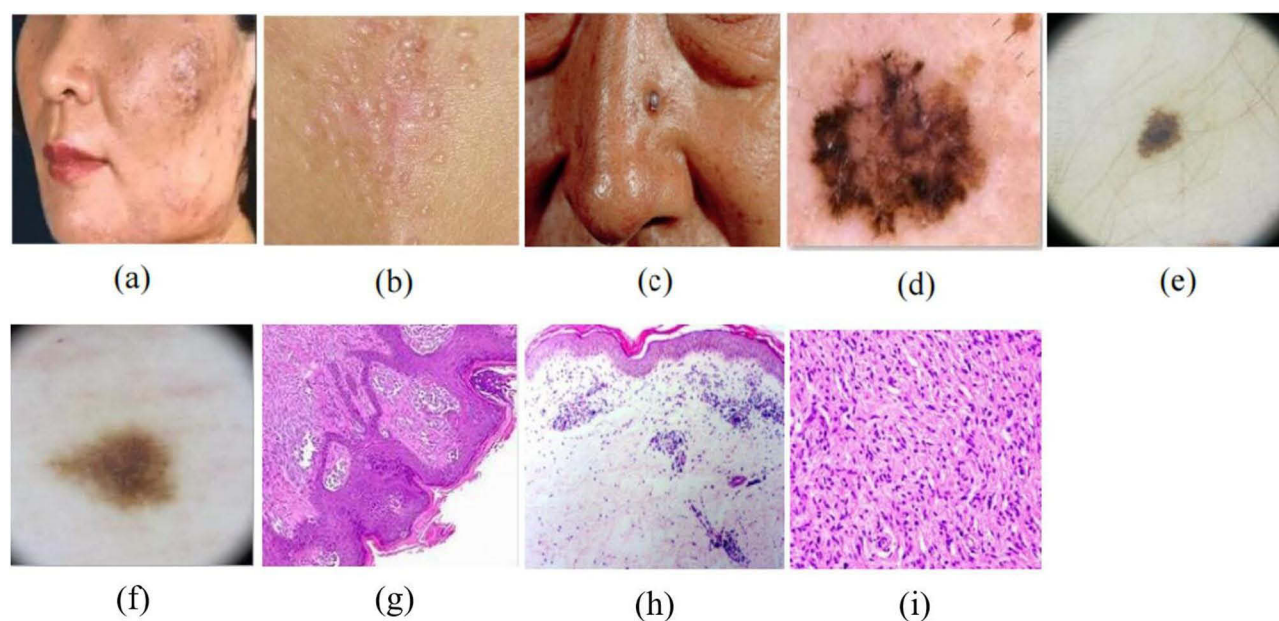

Figure 1 Comparison of clinical skin images (**a–c**), dermoscopic images (**d–f**), and tissue-biopsy pathology images (**g–i**). The gold standard for skin disease diagnosis is based on tissue-biopsy pathology images, but high-quality equipment and testing techniques are required. Dermoscopic images are obtained using a microscope, which can only test lesions in block or dot shape. Clinical images are influenced by the angle of photography and the intensity of light, and are almost always obtained under different lighting conditions and uneven focal lengths of the lesion, resulting in greater external interference.

unlike a pretrained fine-tuning model, in which the classifier uses a previously learned representation and tunes its parameters to maximize the accuracy of new data. The problem with pretrained fine-tuning is that the classifier is most likely to overfit the new data when fewer examples are provided.

State of the Art

Currently, FSL methods are used frequently in early skin disease–screening application tasks such as IoMT[16] and mobile diagnosis platform,[17] FSL methods proven to have satisfactory performance over long-tailed and sample-classification tasks like skin disease classification. Several improved skin lesion–classification methods based on FSL have also been introduced, containing multiperspective improvements on vanilla methods. Liu et al introduced an embedding-based comparison method improved from RelationNet,[18] Zhu et al proposed a novel loss-function calculation named query-relative loss over meta-learning,[19] Li et al and Prabhu et al raised a similar issue of focusing on clustering methods using category judgment in feature maps,[20,21] and Mahajan et al proposed an enhanced convolution strategy, ie, group-equivariant convolutions (G-convolutions) to alleviate classification error brought about by data augmentation, such as rotations and reflections.[22] However, these studies were limited to nonuniversal approaches or few disease classes.

In this work, we propose a multiscale feature-fusion FSL classification approach to skin disease classification, CDD-Net (Context Feature Fusion and Dual Attention Dermatology Net), based on the special texture characteristics of clinical skin images. CDD-Net focuses on feature-level improvement, with the application of the developed extraction and fusion strategy, which makes it possible to utilize almost all the current FSL methods, rather than improvement only in specific tools, such as metric learning–based methods[16,20,21] or meta-learning–based methods.[18,19]

To test our method, we built a large skin disease dataset, Derm104, with 104 classes of skin lesions, which provides a large benchmark for skin disease FSL classification. It merges 1781 self-collected images from a public skin disease atlas website (www.med126.com/pf) and 2702 images from the public dataset SD-198,[23] for a total of 4483 images. For a more stable verification result and to ensure adequate generalization, we ensured that there were at least 20 images per category. Our contributions are as follows.

1. We built a clinical skin disease dataset, Derm104, with 104 classes of skin lesion, providing a reliable benchmark for our research.
2. We propose CDD-Net as a universal feature-level plug-in module in clinical skin image FSL classification fitting the special texture characteristics of skin lesions, which makes it an improvement on almost all the current FSL methods, with accuracy improvement of up to 9.14 percentage points. This approach provides a solving pattern for the scarcity of annotated data problems in clinical skin disease diagnosis.

Related Work and Materials
FSL Methods

In contrast to traditional classification methods, FSL aims to classify instances from undiscovered classes with a small number of labeled samples in each class.[15] Ravi and Larochelle used an LSTM-based meta-learner that captures both short-term knowledge that is specific to a task and long-term knowledge that is shared across all tasks.[24] Using the prototypes for each class, ProtoNet performs classification using a representation-based metric space.[25] The use of FSL in the context of meta-learning has also been investigated by Finn et al,[26] wherein they presented an algorithm for fast adaptation of networks on versatile tasks and showed their effectiveness in one-shot learning tasks. Vinyals et al proposed a network known as matching networks that learns the mapping between a small labeled support set and an unlabeled example.[27] The training process is based on the idea that testing and training settings should be identical. Finn et al continued to investigate meta-learning mixed with imitation learning from a visual demonstration challenge of one-shot learning for a robot.[28] Meta-learning and transfer learning were integrated by Sun et al to develop a hard-task meta-batch scheme, an effective learning curriculum that increases convergence and accuracy.[29]

Typically, each few-shot classification task $T_\square$ corresponds to every training episode in one epoch, and would be randomly selected from a sample dataset, and samples in task $T_\square$ would be divided into support set T_s and a query set T_q.

If the support set T_s contains N classes with K samples in each class, the few-shot classification task is called N-way K-shot. The query set T_q contains the samples from the same classes with the support set T_s. Formally, a few-shot task can be defined as $T=(T_s, T_q)$, where $T_S = \{(x_i^s, y_i^s)\}_{i=1}^{N \times K}$ and $T_q = \{(x_i^q, y_i^q)\}_{i=1}^{Q}$. Given the support set T_s, our goal is to classify and assign the samples in the query set T_q correctly to one of the N classes. This kind of training mechanism based on task training is called metaparadigm or episodic paradigm. Figure 2 shows how episodic training works in skin disease classification.

There are three types of FSL classification:, pretrained fine-tuning–based methods, meta-learning–based methods, and metric learning–based methods.[25–34] In pretrained fine-tuning, a generic training task is used rather than episodic training in the training stage. The entire base class N^{base} is trained to train a classifier:

$$\Gamma(\theta, \omega) = argmin \sum_{i=1}^{N^{base}} L(C_\omega(f_\theta(x_i)), y) \qquad (1)$$

where $\Gamma(\theta, \omega)$ is a parameter of the embedding backbone and the classifier. In the testing stage, several episodic training tasks T were built and sampled from the domain of the novel class N^{novel}. The support set T_s will fine-tune the trained $\Gamma(\theta, \omega)$, and the query set T_q is tested in the classifier.

In meta-learning–based methods, episodic training is employed in both the training stage and testing stage. A meta-learner is added and the network training is processed in a two-step optimization between the base learner and meta-learner. The first step is called base learning or inner loop, which optimizes the base learner with the support set T_s, as in MAML,[26] and the gradient is updated:

$$\Gamma_m = \Gamma_{m-1} - \alpha \nabla_{\Gamma_{m-1}} L_S(F_{\Gamma_{m-1}}) \qquad (2)$$

where $F_\Gamma = f_\theta \circ C_\omega$, α and m are the learning rate and the iteration number in the inner loop, respectively. The second step is meta-fine-tuning or outer loop, which optimizes the meta-learner with query set T_q, and a similar gradient is updated:

$$\Gamma = \Gamma - \beta \nabla_\Gamma L_Q(F_{\Gamma_m}) \qquad (3)$$

The parameter of the model is truly updated over the previous parameter Γ rather than Γ_m by using the query set T_q. In this way, the meta-learner is expected to learn a type of across-task meta-knowledge that can be used for fast adaptation to novel tasks.

In metric learning–based methods, episodic training is employed in the training stage, but not in the testing stage. Instead, the testing stage will directly compare the similarities between the samples in the support set T_s and the query set T_q, as part of a single feed-forward pass mechanism. Every sample is encoded into the latent representation space, as in ProtoNet,[25] and samples are encoded as prototype C_i, $C_i = \frac{1}{K} \sum_{j=1}^{K} f_\theta(x_j)$, where K denotes K-shot in the support set. The prototype of the query sample X^Q computes the distance with C^i, and the predicted posterior probability distributions are:
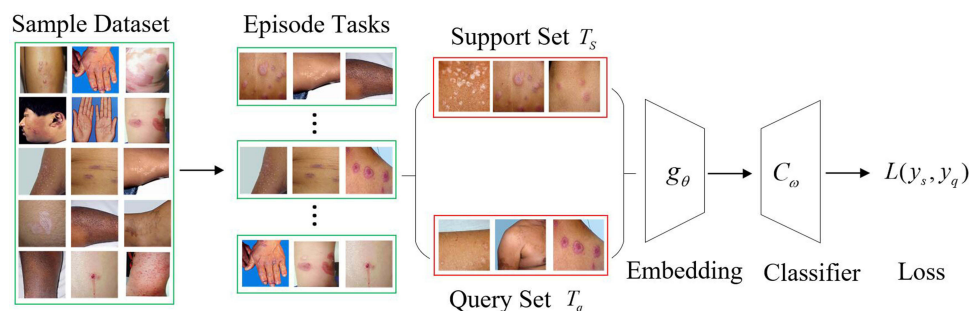

Figure 2 Episodic training mechanism in skin disease classification. A series of episode tasks are built from the sample dataset and divided into support set (T_s) and a query set (T_q). Convolutional neural network–based embedding backbone and classifier noted as f_θ and C_w respectively. Also, the classifier can be integrated with the embedding backbone into a unified network and trained in an end-to-end manner.

$$\rho(y = i|Q) = \frac{\exp(-D(f_\theta(x_Q), C_i))}{\sum_{j=1}^{C} \exp(-D(f_\theta(x_Q), C_i))} \tag{4}$$

where D (.) denotes the Euclidean distance between the prototypes. In the training stage, the standard cross-entropy loss was employed to train the entire model. In addition, during the testing stage, the nearest-neighbor classifier (1–NN) can be conveniently used for prediction.

Dataset

In order to support the application of computer-aided diagnosis in dermatology, several public datasets of skin diseases were created as a benchmark for measuring the effectiveness of automatic diagnosis. The Dermofit dataset is provided by researchers at the University of Edinburgh in Scotland.[35] This dataset is of high quality and is widely used by researchers, but it is not freely available. Dermofit includes 10 different types of skin diseases, but there are only roughly 1300 photos in total. The International Skin Imaging Collaborative has collected a large dataset of publicly accessible dermoscopic images,[36] which was released in 2016 for a publicly available benchmark competition for dermoscopic image analysis. MoleMap is a skin dataset consisting of three cancer categories and 22 benign categories.[37] The cancer categories are melanoma, basal cell carcinoma, and squamous cell carcinoma. Each lesion image consists of a dermoscopic image and a clinical image that was captured 10 cm from the lesion. The Asan dataset was obtained from the Dermatology Department of Asan Hospital.[38] It contains 12 dermatological disease categories within Asian subjects. A collection comprising 1276 of these photos is available for downloading for research purposes. SD-198 is a publicly available dataset of clinical skin images collected by Sun et al. It comprises 6584 images range in size, color, shape, and structure across 198 categories.[23]

All of the abovementioned datasets have their own disadvantages in clinical skin disease classification. The International Skin Imaging Collaborative dataset mainly contains dermoscopic images, which are costlier and more difficult to obtain from clinical images.[36] The Dermofit, MoleMap, and Asan datasets contain fewer categories (<25) and do not cover enough rare categories.[35,37,38] Compared with other datasets, SD-198 is a more proper dataset,[23] but more than half of its collected categories contain fewer than 20 images, which may cause overfitting during the training of a deep-learning network. Another problem is that some major species and subspecies are not classified accurately, such as general psoriasis, pustular psoriasis, and nail psoriasis, which may cause category confusion.

We present a new clinical skin disease dataset as a benchmark for clinical skin disease classification, Derm104, which covers sufficient rare diseases and contains easily obtainable clinical images. Derm104 has high diversity, multiple categories, wide coverage, and reliable samples and benchmark experiments can effectively evaluate the reliability of various FSL methods in clinical skin disease classification.

1. Annotation and Scale: Derm104 contains 4483 clinical images covering 104 types of skin disease and is merged from two parts. The first part comprised 1781 images, all of which were collected by professional dermatologists from the public dermatology dataset on www.med.126.com/pf. The second part contains 2702 images from the public dataset SD-198. Both parts of the dataset are public and freely available to researchers, and the patients involved in the database have given approval for their images to be used. To prevent ambiguity and ensure the accuracy of classification between subspecies, we chose some of the images from SD-198, and every category ground truth was verified by professional dermatologists. Images per class vary from 20 to 141 to ensure variance stability.

2. Diversity: All images are from a real scene with variance in color, exposure, illumination, and level of detail collected in a variety of environments. Also, skin lesion images include: eczema, psoriasis, acne, lichens, prurigo, alopecia areata, urticaria, scabies, nervus, erythema, pityriasis, vercurra, bacterial skin diseases, and melanoma skin cancer, covered more than 90% of the incidence of all skin diseases.

3. Fine Granularity: Our dataset also focuses on fine-grained disease classes, such as many subspecies that have tiny interclass variance and are difficult to classify. For example, erythema, pityriasis, and lichen diseases can be classified into many subcategories that are difficult to distinguish. Few subspecies are included in existing datasets,[35–38] so Derm104 focuses more on this. Some of our fine-grained images are shown in Figure 3.
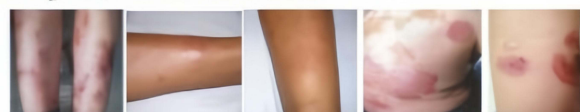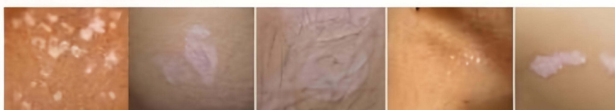
Figure 3 Fine-grained classified images. Collected subspecies in erythema (**a**), lichen (**b**), and pityriasis (**c**) as examples. Images in such subspecies have tiny interclass variance and significant intraclass variance, causing challenges in correctly diagnosing similar dermatological diseases.

Methods

The classification of clinical skin lesions emphasizes lesion location and examination of textural features compared with natural scene images. For example, locations such as the cheek, nose, or eye influence the determination of facial skin diseases, and texture features such as line shape and large flake area cannot be easily focused on, rather than dot or block shapes. However, normal classification methods may not work efficiently. In order to extract specific deep features in skin lesions, we propose CDD-Net as an enhanced solution for clinical skin disease classification. Current FSL methods for skin disease classification are limited to individual approaches,[18–22] but CDD-Net has highly versatile performance that can be used as a plug-in module in both metric learning–based, pretrained fine-tuning–based and meta-learning–based FSL methods.

CDD-Net is composed of a context feature–fusion module and a dual-attention module for different stages of feature processing. The context feature–fusion module localizes the discriminatory texture details of skin lesions by integrating features from different layers. The dual-attention module highlights discriminative regions from channel-wise and pixel-wise representations based on weight vectors and restrains the contributions of irrelevant areas. In the following section, we describe the entire framework of the proposed CDD-Net. The overall flowchart is given in Figure 4.
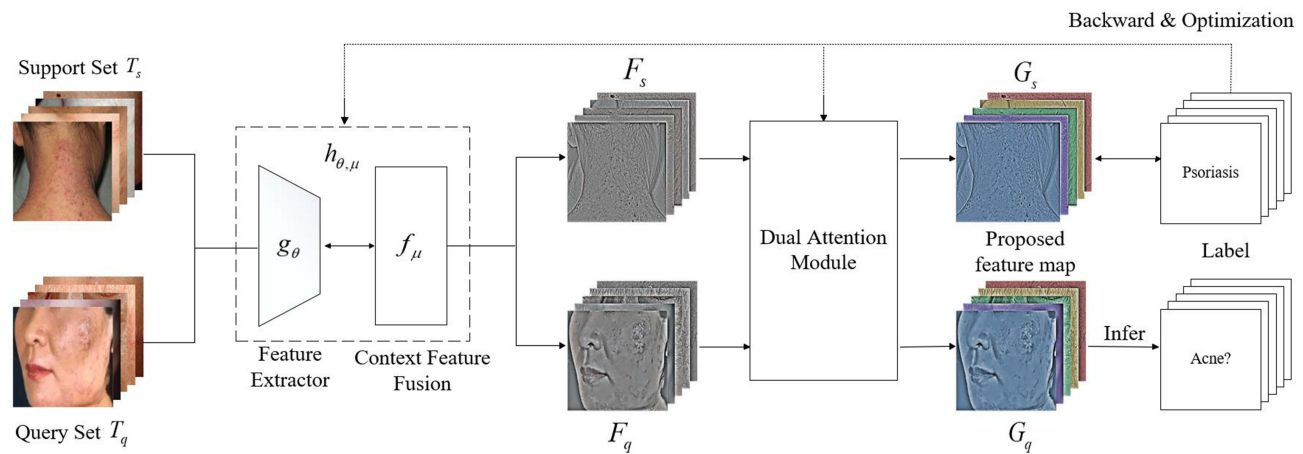
1012

DovePress

Clinical, Cosmetic and Investigational Dermatology 2024:17

Figure 4 Overview of our approach. CDD-Net consists of two submodules: context-feature fusion and dual attention. An episode batch comprising support set T_s and query set T_q will be processed by the feature extractor $h_{\theta,\mu}$, which is combined by the CNN-based network g_θ and proposed context feature–fusion module f_μ. g_θ can be supported with fusion module f_μ and utilizes hidden texture information. After that, the dual-attention module will work on a feature map of support set F^s and query set F^q. Feature maps will be categorized into G^s and G^q, with channel-wise and pixel-wise discriminant information. **During episode training, the support set needs to learn with the labels and propose backward optimization into the context feature –fusion module and dual -attention module.** The optimized modules will help the query set to infer new classes in episode testing.

Context Feature–Fusion Module

In recent years, the idea of feature fusion has been applied in many visual tasks. Specifically, Long et al combined deep semantic features with shallow appearance features to produce accurate and detailed segmentations.[39] Manzoor et al effectively fused deep features extracted from different convolution layers for high-resolution dermoscopic skin image classification.[40] In this paper, we introduce a context feature–fusion mechanism to exploit the strong complementary and correlated information among different hierarchical layers for clinical skin image classification.

Considering that different layers may have feature maps of different sizes, we use the bottleneck module to ensure that they have the same spectral dimensionality before feature fusion. Such an operation is applicable to convolution layer-based neural networks, and thus we can utilize our module within CNN networks such as ResNet and others. For instance, in ResNet12, we assume that F_1, F_2, F_3, and F_4 refer to the outputs from low to higher-level layers that have 64, 160, 320, and 640 feature maps, respectively. Then, we can use 640 kernels with of size 1×1 to convolute them. With such a convolution operation, the number of feature maps from F_1 to F_4 becomes 640. Finally, feature fusion can be performed easily using element-wise summation. The entire process can be represented as:

$$F_z = \frac{1}{n}\sum_{L=1}^{n} B_L(F_L) \tag{5}$$

where F_Z represents the fused features, F_L represents features from different convolution layers, and B_L is a bottleneck module corresponding to F_L. Conversely, F_Z can also be represented as the result of sample image x_i processed through our feature extractor $h_{\theta,\mu}$, where $h_{\theta,\mu}$ is combined by convolution network g_θ and feature-fusion module f_μ:

$$\begin{aligned} h_{\theta,\mu} &= g_\theta \circ f_\mu \\ F_z &= h_{\theta,\mu}(x_i) \end{aligned} \tag{6}$$

Finally, we use pooling as the global averaging function. The entire context feature–fusion module is illustrated in Figure 5.

Dual-Attention Module

Introducing an attention mechanism to complete the adaptive weight adjustment is helpful in highlighting the representative discriminative regions and restraining the contributions of irrelevant backgrounds. Ma et al used a squeeze-and-excitation module to enhance channel weight on new feature maps via pooling and full connection layers, elevating
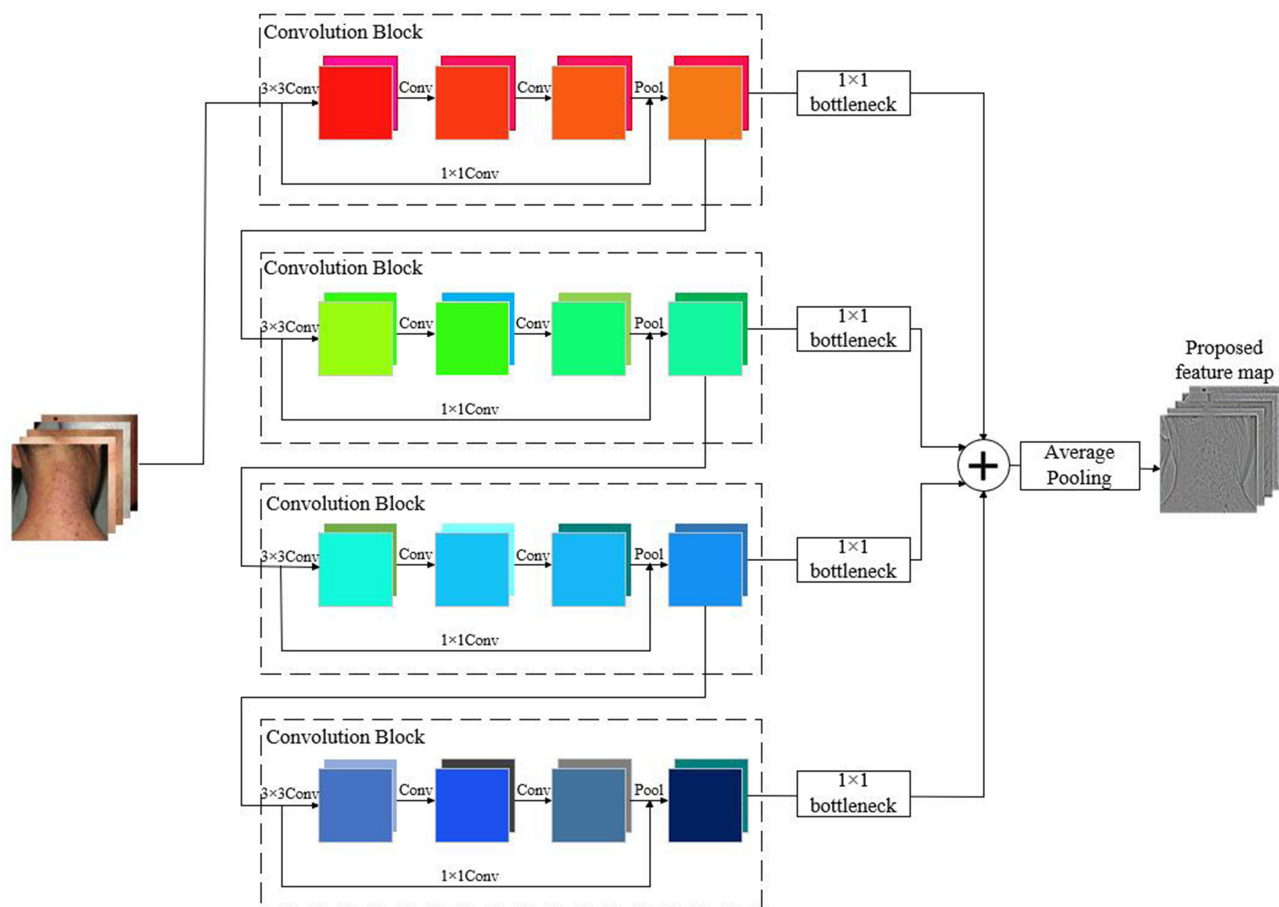
Figure 5 Schematic illustration of context feature–fusion module. A traditional convolutional neural network would pass original images through a deep network consisting of several convolutional blocks. Different hierarchy blocks would output feature maps of various dimensions incorporating different receptive fields, resolutions, and detailed information. We propose average-pooling our processed feature maps. For convenience, the batch normalization, activation layers, and dimension-matching operations are not given.

segmentation efficiency for brain tumor imags.[41] Space weight focuses on the location of the area of interest. Wei et al proposed a spatial pyramid pooling module for lesion segmentation of dermoscopic images.[42]

We propose a dual-attention mechanism to enhance the ability to exploit features of interest based on sensitivity scores. For each episode of the support set and query set, we define the channel-wise and pixel-wise sensitivity score to quantify the representativeness. Prior to explaining scores, we first define feature maps F of the support and query instance as follows:

$$F_{i,j}^s = h_{\theta,\mu}\left(x_{i,j}^s\right)$$
$$F^q = h_{\theta,\mu}(x^q)$$

(7)

where $x_{i,j}^s$ is the j-th instance of the i-th class in the support set, X_q is the query instance, and $h_{\theta,\mu}$ is a feature extractor supported by the proposed context-feature fusion in (Equation 6). Specifically, we utilize a prototype as the representative of each class in the support set.[25] The prototype is defined as:

$$F_i^p = \frac{1}{k}\sum_{j=1}^{k} F_{i,j}^s$$

(8)

where k is the instance size for each support class (ie, k-shot). Based on this, $m_{i,c}^p \in \mathbb{R}^{H \times W}$ and $n_{i,x}^p \in \mathbb{R}^C$ mean the c-th channel and x-th pixel of the prototype feature map $F_i^p \in \mathbb{R}^{C \times H \times W}$, respectively, and we compute the mean features in channel and spatial dimensions to represent the standard condition in each class:

$$M^p = \frac{1}{C} \sum_{z=1}^{C} m_{i,z}^p$$

$$N^p = \frac{1}{H \times W} \sum_{z=1}^{H \times W} n_{i,z}^p \tag{9}$$

Furthermore, we obtain the channel-wise and pixel-wise sensitivity scores for each channel and pixel. Note that we compute scores in the feature maps of support images rather than prototypes, where $F_{i,j}^s \in \mathbb{R}^{C \times H \times W}$ indicates the feature map on the j-th instance of the i-th class in the support set. Similarly to what we have calculated on the aforementioned prototype feature, $m_{i,j,c}^s \in \mathbb{R}^{H \times W}$ and $n_{i,j,x}^s \in \mathbb{R}^C$ refer to the c-th channel and x-th pixel of $F_{i,j}^s$, respectively:

$$w_{i,j,c}^{channel} = \frac{1}{H \times W} \| m_{i,j,c}^s - M^p \|^2$$

$$w_{i,j,x}^{pixel} = \frac{1}{C} \| n_{i,j,x}^s - N^p \|^2 \tag{10}$$

Intuitively, sensitive scores represent the discriminative information between the mean standard and region of interest, but cannot finally determine how much attention should be paid to this. In order to choose efficient weights and improve the generalization, we introduce the fully connected block (FCB) to realign the weight vectors. The architecture of the FCB is presented in Table 1.

Notes: S, input weight that would be realigned. When FCB is required for adaption of channel-wise depiction, S is equal to channel size C, while FCB requires pixel-wise depiction and S is equal to layer size H×W.

The sensitivity scores in a single-feature map can be combined into weight vectors $w_{i,j}^{\text{channel}} \in \mathbb{R}^C$ and $w_{i,j}^{pixel} \in \mathbb{R}^{H \times W}$, and they would be transformed through a fully connected block corresponding to the discriminative information from the perspectives of channel and pixel, respectively, and processed thus:

$$v_{i,j}^{channel} = H^{channel}\left(w_{i,j}^{channel}\right) + noise$$

$$v_{i,j}^{pixel} = H^{pixel}\left(w_{i,j}^{pixel}\right) + noise \tag{11}$$

where H represents FCBs that adapt to the different weight vectors. In the episode-training stage, we add random noise between −0.2 and 0.2 to the weight vector $v_{i,j}$ to prevent overfitting. Upon obtaining channel- and pixel-discrimination weight vectors, we scale them with the original feature map correspondingly. Finally, we construct a linear combination with a balancing hyperparameter α:

$$G_{i,j}^s = \alpha \left(v_{i,j}^{channel} \odot F_{i,j}^s\right) + (1 - \alpha)\left(v_{i,j}^{pixel} \odot F_{i,j}^s\right), \alpha \in [0, 1] \tag{12}$$

In a similar vein, we can also obtain the processed feature map $G^q \in \mathbb{R}^{C \times H \times W}$ from the query image. The difference is that we use its own feature map rather than the prototype in the support set. The entire dual-attention module is shown in Figure 6.

Table 1 Architecture of the fully connected block.Fully Connnection Block

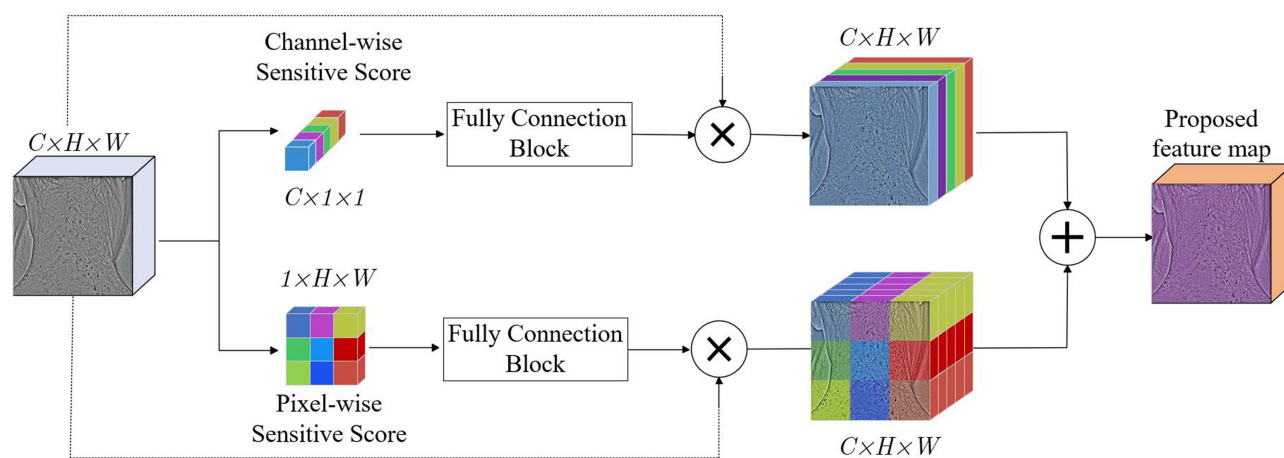| Layer	Output size
Input	S
Fully connnected layer	2S
Batch normalization	2S
ReLU	2S
Fully connnected layer	S
1+Tanh	S

Figure 6 Schematic illustration of dual-attention module. The different colors in the cubes indicate different sensitive scores in channels and pixels. In addition, calculations of sensitive scores for individual feature maps in the support set and query set, channel-wise and pixel-wise, are independent of one another. The final combination from the perspectives proposed is dependent on hyperparameter $\alpha \in [0,1]$ (Equation 12), which is omitted in the illustration.

Results

In this section, we evaluate CDD-Net on the proposed Derm104 benchmark. Derm104 consists of 4483 clinical images covering 104 types of skin diseases. Derm104 has detailed and reliable labels, enough categories, high diversity, and fine granularity, and thus we propose it as our benchmark for clinical skin disease FSL classification. In sum, 62 categories were set as the training set, with 21 categories as the validation set and another 21 categories as the testing set.

We adopted common protocols from recent FSL research on skin disease.[18–22] The entire training procedure was completed on the GPU of an Nvidia GeForce 1080 Ti. The input image size was set to 84×84×3, and lightweight backbones of Conv64F and ResNet12 were used. ResNet12 produced a feature map with an output size of 640×5×5, whereas Conv64F provided a size of 64×5×5. Prior to inputting the photos into the convolution network, common data-augmentation methods, such as center crop and horizontal flip, were employed. Note that skin lesion classification is color-sensitive, so color jitter is not used.[43] Parameter α (Equation 12) was fixed at 0.5.

Following common FSL methods,[15,24,27] we set five-way one-shot and five-way five-shot for the episode-training paradigm and 10 query images per new class for result evaluation. Episodes of training, validation, and testing were set to 100 per epoch, and the entire learning procedure consisted of 100 epochs. We set a learning rate of 1^{-3} for the Adam optimizer. As for some approaches that require another learning rate in their inner circle, we set rates of 1^{-3} for ANIL and 1^{-2} for baseline++ and MAML. In the fully convoluted block of the dual-attention module, we set a rate of 1^{-4}. All these learning rates were step-scheduled to 0.1 after 70 epochs.

To verify the high adaptability of CDD-Net, we applied it to various FSL methods from different approaches: the pretrained fine-tuning–based methods baseline++ and RFS-Distill, meta-learning–based methods MAML, ANIL, metric learning–based methods ProtoNet, DN4, EPNet, and Neg-Cosine.[25–27,30–34] The classification accuracy was within the 95% confidence interval. Tables 2 and 3 show the performance of our proposed CDD-Net on five-way one-shot and five-way five-shot clinical skin disease classification, respectively.

As shown in Tables 2 and 3, we clearly validated the impressive performance, generalization ability, and robustness of the proposed CDD-Net. It is versatile for pretrained fine-tuning–based, meta-learning–based, and metric learning–based FSL methods, with accuracy improvement of up to 9.14 percentage points. Since CDD-Net has better performance on metric learning–based approaches, we conducted an ablation study on ProtoNet, DN4, EPNet, and Neg-Cosine with the modified CDD-Net and report the efficiency of its submodules.

We additionally explored the effects of the dual-attention module in CDD-Net. In Table 4, the first row shows the results without any modules and the third row shows the accuracy after adding dual attention. We found that perceiving the relative location of skin lesion and screening task-related regions is as important as emphasizing the channels that

Table 2 Performance of CDD-Net on five-way one-shot classification with comparison of various methods

Method	Embedding	Type	One-shot	With CDD-Net
Baseline++[30]	Conv64F	Pretrained fine-tuning	51.19	52.60
RFS-Distill[31]	Conv64F	Pretrained fine-tuning	52.36	58.40
MAML[26]	Conv64F	Meta	46.32	47.59
ANIL[27]	Conv64F	Meta	49.63	51.98
ProtoNet[25]	Conv64F	Metric	42.53	49.08
DN4[32]	Conv64F	Metric	54.30*	59.46*
EPNet[33]	Conv64F	Metric	53.68	58.79
Neg-Cosine[34]	Conv64F	Metric	53.02	56.62
Baseline++	ResNet12	Pretrained fine-tuning	53.45	51.71
RFS-Distill	ResNet12	Pretrained fine-tuning	53.53*	57.52
ANIL	ResNet12	Meta	47.35	51.31
ProtoNet	ResNet12	Metric	52.06	59.39*
DN4	ResNet12	Metric	50.09	56.94
EPNet	ResNet12	Metric	52.36	59.63
Neg-Cosine	ResNet12	Metric	52.48	54.67

Notes: *Methods with outstanding performance. In vanilla approaches, DN4 performs best in Conv64 embedding and RFS-Distill performs best in ResNet12 embedding. In optimized CDD approaches, DN4 performs best in Conv64 embedding and ProtoNet performs best in ResNet12 embedding.

Table 3 Performance of CDD-Net on five-way five-shot classification with comparison of various methods

Method	Embedding	Type	Five-shot	With CDD-Net
Baseline++[30]	Conv64F	Pretrained fine-tuning	63.08	66.75
RFS-Distill[31]	Conv64F	Pretrained fine-tuning	68.29*	73.01
MAML[26]	Conv64F	Meta	55.50	57.74
ANIL[27]	Conv64F	Meta	53.61	54.54
ProtoNet[25]	Conv64F	Metric	60.75	68.05
DN4[32]	Conv64F	Metric	68.10	73.78*
EPNet[33]	Conv64F	Metric	66.32	71.73
Neg-Cosine[34]	Conv64F	Metric	65.53	70.08
Baseline++	ResNet12	Pretrained fine-tuning	64.60	67.39
RFS-Distill	ResNet12	Pretrained fine-tuning	69.04	73.71
ANIL	ResNet12	Meta	49.02	52.91
ProtoNet	ResNet12	Metric	64.14	73.28
DN4	ResNet12	Metric	73.30*	78.58*
EPNet	ResNet12	Metric	71.38	76.37
Neg-Cosine	ResNet12	Metric	70.53	74.64

Notes: *Methods with outstanding performance. In vanilla approaches, RFS-Distill performs best in Conv64 embedding and DN4 performs best in ResNet12 embedding. In optimized CDD approaches, DN4 both performs best in Conv64 embedding and ResNet12 embedding.

extract key features. A single attention model is not sufficiently effective, and the second row in Table 4 shows the influence of using only a channel-wise sensitivity score in the classification.

Furthermore, we studied the fusion strategy in the context feature–fusion module. The first row in Table 5 shows the results without any feature-fusion operations, as well as the third row in Table 4. The results of concatenative and additive fusion are shown in the second and third rows, respectively. We suppose that the concatenation of different layers causes the channel dimension to overinflate (up to 2560 in ResNet12), which may cause overfitting and dimension

Table 4 Performance of different attention-selection strategies in dual-attention module

Method	Embedding	Type	Ablation strategy	One-shot	Five-shot
ProtoNet[25]	ResNet12	Metric	Original	52.06	64.14
DN4[32]				50.09	73.30*
EPNet[34]				52.36	71.38
Neg-Cosine[44]				52.48*	70.53
ProtoNet	ResNet12	Metric	Channel-wise attention	53.63	65.64
DN4				51.12	74.31*
EPNet				54.05*	72.39
Neg-Cosine				52.96	71.48
ProtoNet	ResNet12	Metric	Dual attention	55.36	69.46
DN4				54.62	75.37*
EPNet				55.83*	73.44
Neg-Cosine				53.80	72.51

Notes: *Methods with outstanding performance. In original approaches, Neg-Cosine performs best in one-shot and DN4 performs best in five-shot. In channel-wise attention approaches, EPNet performs best in one-shot and DN4 performs best in five-shot. In dual-attention approaches, EPNet performs best in one-shot and DN4 performs best in five-shot.

Table 5 Performance of different fusion strategies in context feature–fusion module

Method	Embedding	Type	Ablation strategy	One-shot	Five-shot
ProtoNet[25]	ResNet12	Metric	W/O feature fusion	55.36	69.46
DN4[32]				54.62	75.37*
EPNet[34]				55.83*	73.44
Neg-Cosine[44]				53.80	72.51
ProtoNet	ResNet12	Metric	Concatenative fusion	56.29	69.88
DN4				54.13	76.35*
EPNet				56.42*	74.40
Neg-Cosine				53.45	73.38
ProtoNet	ResNet12	Metric	Additive fusion	59.39	73.28
DN4				56.94	78.58*
EPNet				59.63*	76.37
Neg-Cosine				54.67	74.64

Notes: *Methods with outstanding performance. In W/O feature-fusion approaches, EPNet performs best in one-shot and DN4 performs best in five-shot. In concatenative fusion approaches, EPNet performs best in one-shot and DN4 performs best in five-shot. In additive fusion approaches, EPNet performs best in one-shot and DN4 performs best in five-shot.

disaster. The bottleneck convolution in the last layer would lose critical gradient information. Instead, additive fusion using average pooling is an easy and reliable approach.

In order to obtain a better measure of the detailed improvement of CDD-Net, for each class of clinical skin disease classification, we add a comparison experiment between DN4 with CDD-Net and several vanilla networks relying on ResNet12 embedding. We demonstrate the result of final epoch running on the testing set comprising 21 categories. This epoch contains 100 episodes of five-way five-shot classification, ie, 50 query images in five ways per episode, for a total of 5000 queries randomly selected from the testing set. We have chosen several methods that perform better than the aforementioned results, including the metric learning method DN4, EPNet, and the pretrained fine-tuning–based method RFS-Distll for further demonstration. The comparison of confusion matrices for query images between vanilla DN4 and CDD-Net is illustrated in Figures 7 and 8.

To evaluate the performance of our proposed method, several evaluation metrics are employed: accuracy ($\frac{TP+TN}{TP+TN+FP+FN}$), precision ($\frac{TP}{TP+FP}$), recall ($\frac{TP}{TP+FN}$), and F_1 score ($2 \times \frac{Precision \times Recall}{Precision + Recall}$). These evaluation metrics are defined

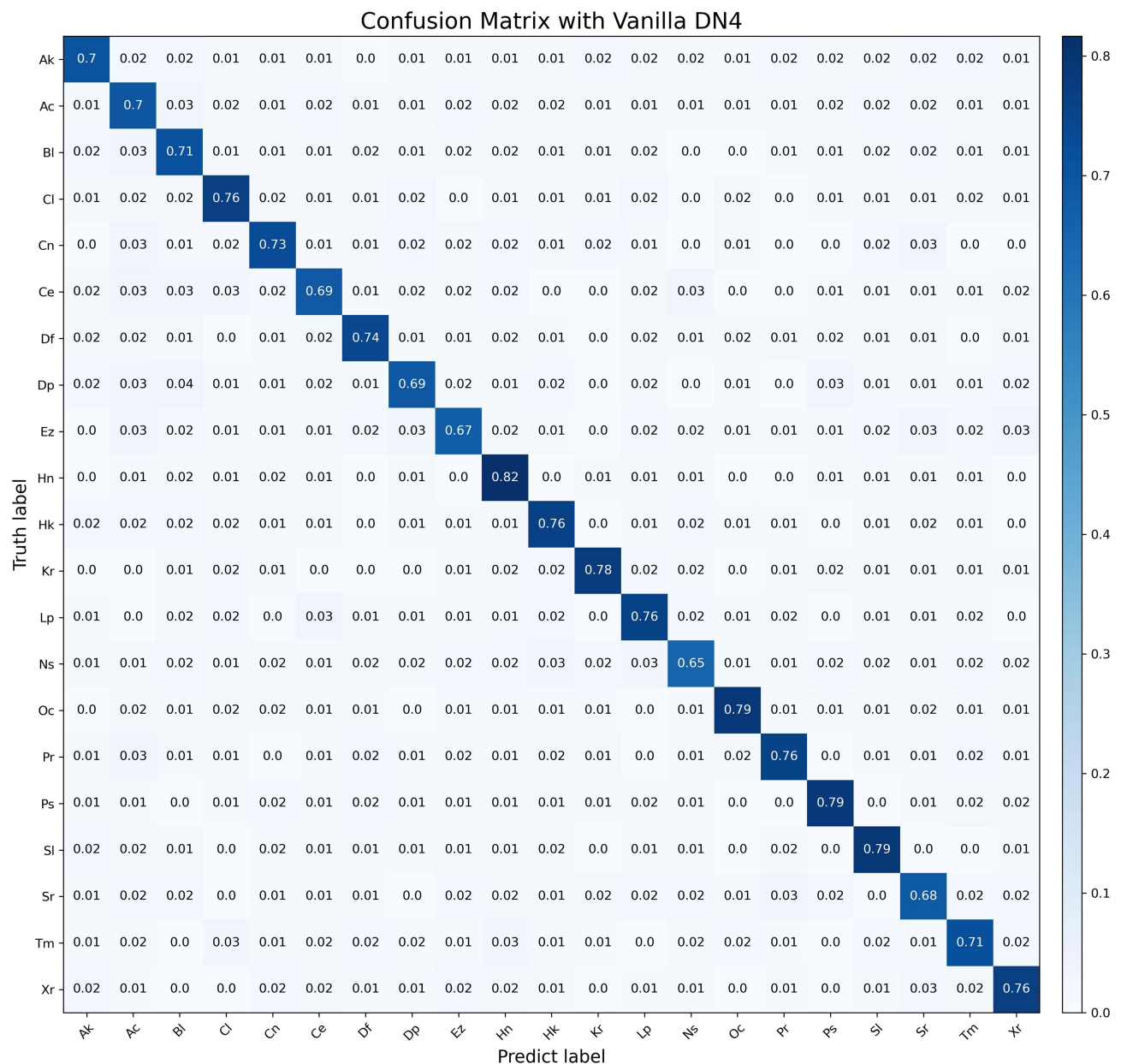## Confusion Matrix with Vanilla DN4

Figure 7 Confusion matrix with vanilla DN4.

based on confusion matrix in which TP, TN, FP, and FN refer to true positive (both real and predicted classes are true), true negative (both real and predicted classes are false), false positive (the actual class of the data is false, while the predicted one is true), and false negative (the actual class of the data is true, while the predicted one is false), respectively. Accuracy is the number of correct predictions over the total number of predictions, also called the classification rate. The number of predicted positive classes that belong to the actual positive class is calculated by recision, while recall determines the true-positive rate. F_1 score measures the testing result according to precision and recall. The comparison of CDD-Net with vanilla DN4, EPNet, and RFS-Distill is illustrated in Figures 9–11.

The comparison of complex matrix and the precision, recall, and F_1 scores focuses on the detailed performance of different classification approaches on different disease categories. The preceding experiment shows that our proposed CDD-Net performs best in most of the testing categories, especially in Cl (cellulitis) and Hn (halo nevus), with a lead nearing 10 percentage points over DN4, but it does not show enough advantage in Dp (dilated pore of Winer), Ns (nevus sebaceous of Jadassohn), or Oc (Onychomycosis).
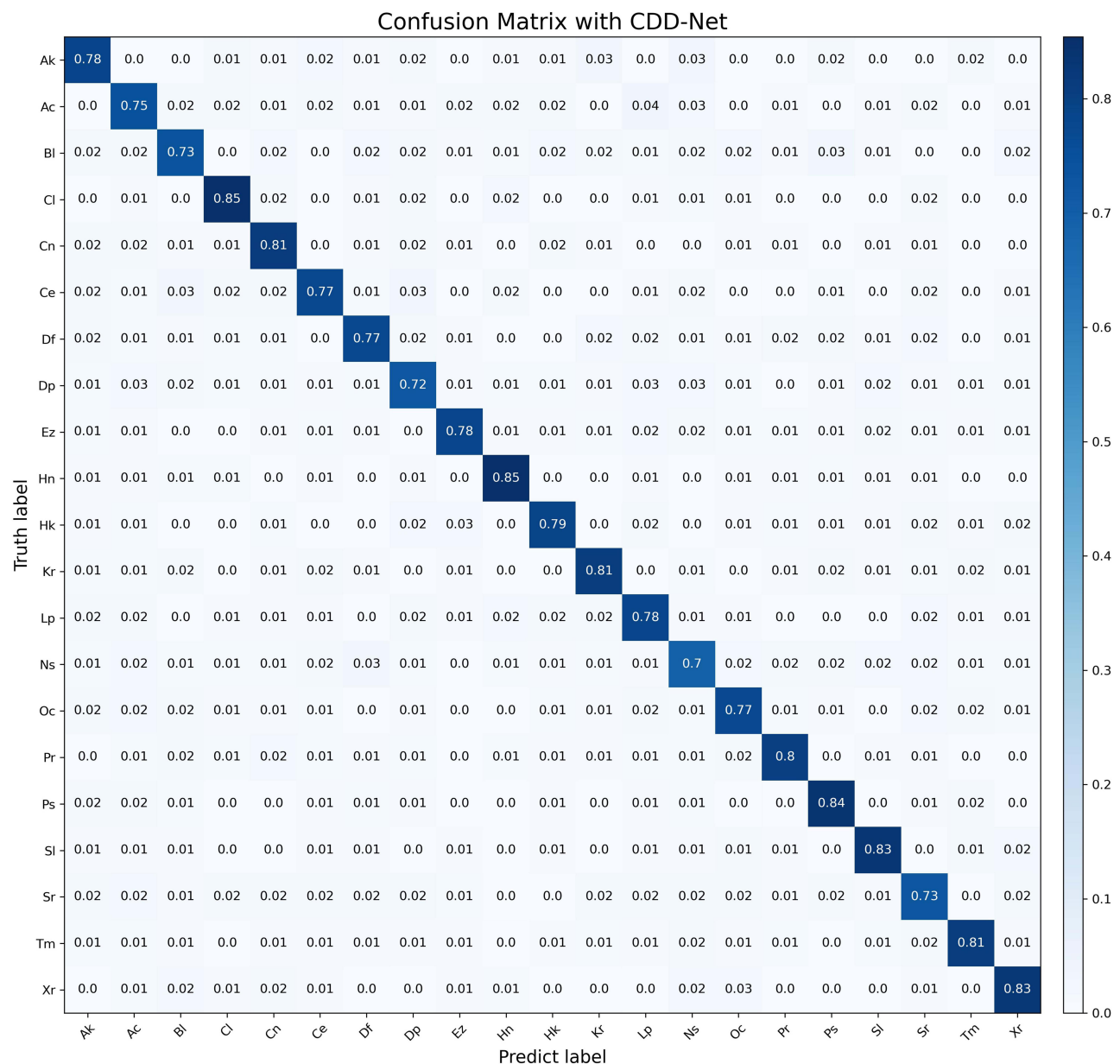
Figure 8 Confusion matrix with CDD-Net.

Top-1 accuracy is the accuracy of the top-ranked category that matches the actual results, while Top-5 accuracy is the accuracy of the top five categories containing actual results. Macro-averaging and micro-averaging methods are also widely used in judging multiclassification tasks. Macro-averaging counts the indicator values for each class and then averages them arithmetically across all classes: $MacroPrecision = \frac{1}{N}\sum_{i=1}^{N} Precision(i)$, $MacroRecall = \frac{1}{N}\sum_{i=1}^{N} Recall(i)$, $MacroF1 = \frac{1}{N}\sum_{i=1}^{N} F1(i)$. Micro-averaging calculates the corresponding metrics based on a global confusion matrix: $MicroPrecision = MicroRecall = MicroF1 = Accuracy = \sum_{i=1}^{N} \frac{TP(i)}{TP(i)+FP(i)}$. The comparison of different methods on several evaluation metrics is demonstrated in Table 6.
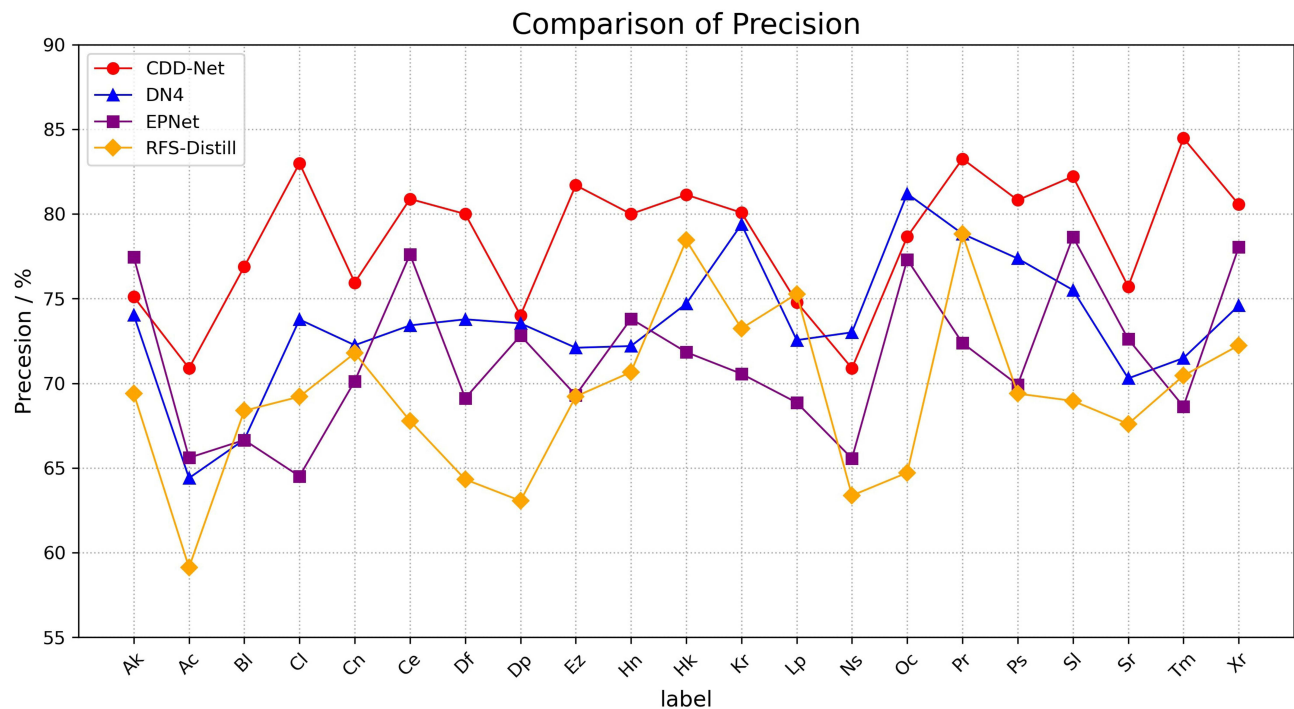
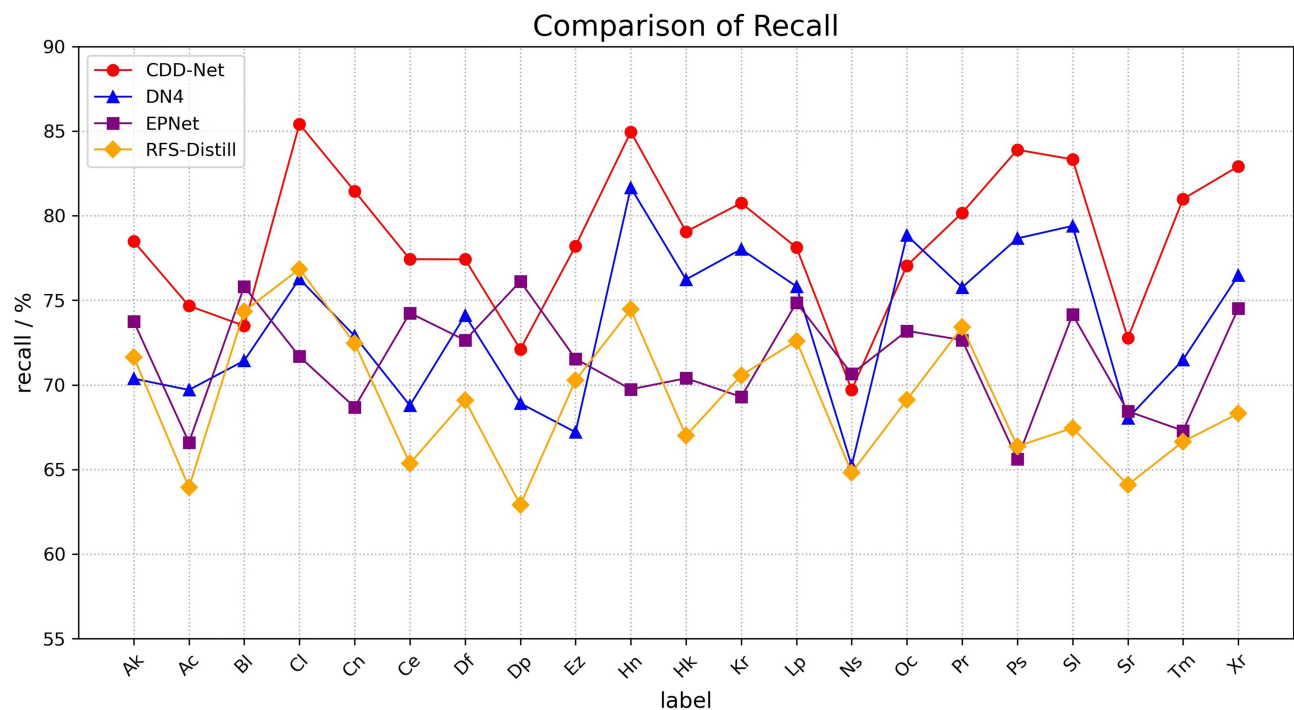**Figure 9** Comparison of precision in CDD-Net, vanilla DN4, EPNet, and RFS-Distill.

Figure 10 Comparison of recall in CDD-Net, vanilla DN4, EPNet, and RFS-Distill.

Discussion

Compared with traditional classification methods, FSL classification **performs better in accommodating** unseen novel classes[15] and has demonstrated great suitability and superiority in the field of clinical dermatology diagnostics,[16,17] since skin disease–classification tasks have have large variety and few samples. The classification accuracy was within the 95%

Comparison of F1-Score
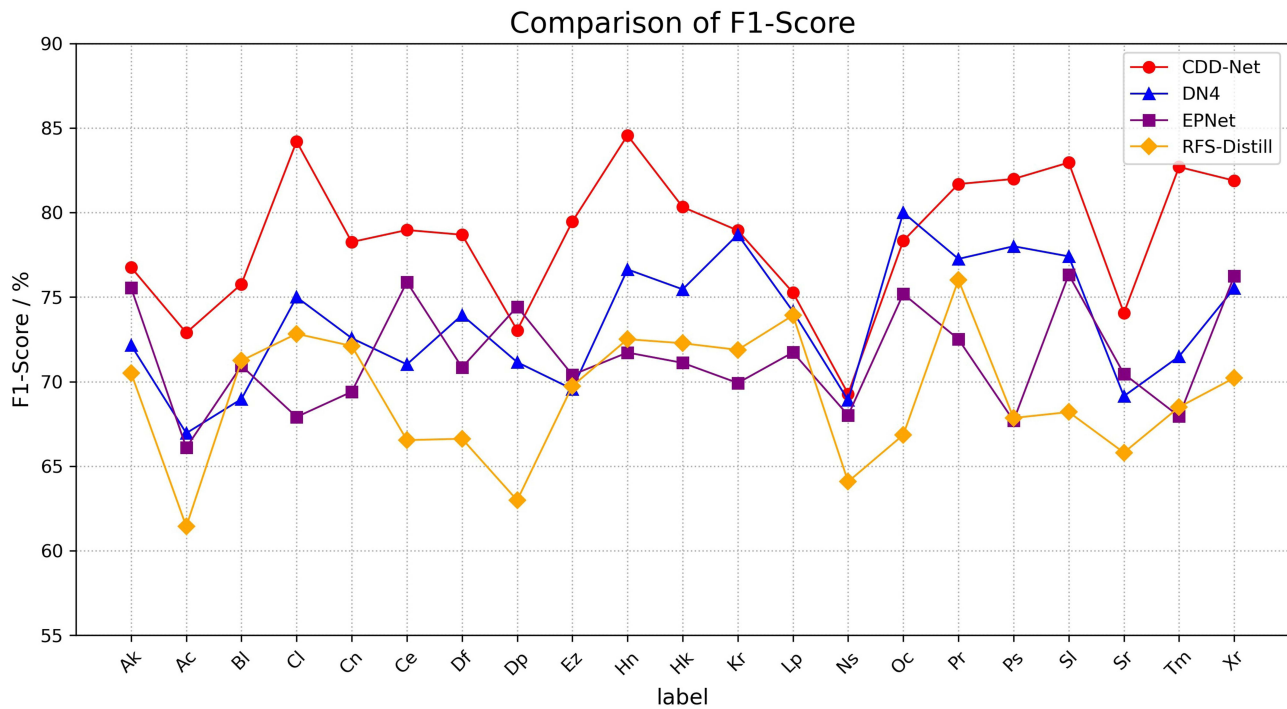
Figure 11 Comparison of F1-Score in CDD-Net, vanilla DN4, EPNet, and RFS-Distill.

confidence interval, and shows our CDD-Net is a versatile and robust development for clinical skin disease FSL classification. Despite FSL approaches using different training paradigms, such as pretrained fine-tuning–based, meta-learning–based, and metric learning–based methods, CDD-Net performs more efficiently and is stable. Each submodule in the multifusion strategy has also been shown to be efficient, **proving that our improved model in this work is feasible and correct**.

Most approaches perform better with more complex backbones, except for meta-learning methods, which are unstable in sensitive networks. Based on the results in Tables 2 and 3, CDD-Net shows better performance on metric learning–based methods than others, especially for the five-way five-shot ResNet12-embedded network: a 4.87 percentage-point improvement versus the pretrained fine-tuning–based method RFS-Distill. We suppose that the reason is that metric learning–based methods have points on the comparison of feature maps between the support set and query set, and our optimization of feature maps fits better to this. However, pretrained fine-tuning–based and meta-learning–based approaches have more dependence on the optimization of the training mechanism. Pretrained fine-tuning–based methods rely more on prior knowledge in the base classes, whereas meta-learning–based methods rely more on justified meta-learner iterations.

It is worth noting that CDD-Net's performance was unsatisfactory when compared to some approaches, including baseline++. We suppose that this is due to our sensitivity scores calculated in the dual-attention module being based on Euclidean distance, whereas baseline++ uses the cosine distance instead of the linear layer as its part of the classifier. Cosine distance focuses more on relative direction, but is not sensitive to absolute distance, and thus the classifier cannot

Table 6 Comparison of different methods on several different evaluation metrics

	Top-1 accuracy (%)	Top-5 accuracy (%)	Macro-precision (%)	Macro-recall (%)	Macro-F1 score (%)
RFS-Distill	69.04	74.73	69.30	69.12	69.14
EPNet	71.38	77.05	71.49	71.51	71.44
DN4	73.30	79.52	73.57	73.58	73.52
CDD-Net	78.58*	84.86*	78.62*	78.68*	78.57*

Notes: *Methods with outstanding performance. Our optimized CDD-Net method performs best in all of the evaluation metrics.
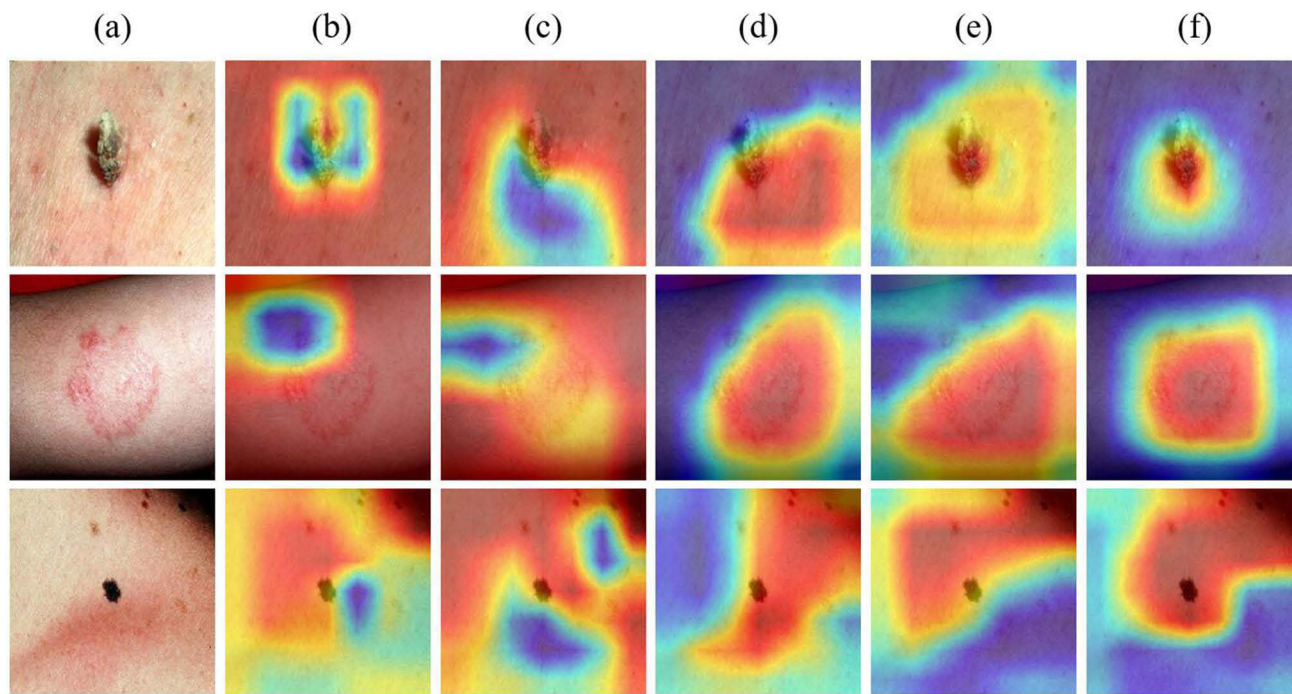
Figure 12 Grad-CAM visualization results: (**a**) lesion images of benign keratosis, tinea corporis, and junction nevus; (**b–e**) ablation on ProtoNet in five-way five-shot mode; (**b**) without any modules in CDD-Net; (**c**) channel-wise attention; (**d**) dual attention; (**e** and **f**) concatenative fusion and additive fusion, respectively, based on dual attention.

take sufficient advantage of our provided template. The result for Neg-Cosine also shows less accuracy improvement than that over other methods. **CDD-Net performance was more robust in five-shot implementations, indicating its reliability with more samples to learn and feedback than standard methods. CDD-Net showed improvement in versatility of 0.93 to 9.14 percentage points over vanilla approaches**, and the best results were for metric learning–based methods.

We chose some skin lesion images from our dataset for Grad-CAM visualization (Figure 12).[44] Three types of representative result image were randomly selected: benign keratosis, tinea corporis, and junction nevus. To show the tendency of the judgment in the neural network, Grad-CAM results for the last layer in the network are demonstrated. Since the ablation results on ProtoNet in five-way five-shot mode showed better improvement, we choose these Grad-CAM results for a more intuitive demonstration. We found that the network's "attention" to the actual lesion becomes more accurate with the improvement in the various modules of the CDD-NET. **Before the dual-attention module was applied,**the tendency of the network to locate a lesion was not sufficiently obvious, and there will be cases in which the zone of interest is confusing and mistaken.

Due to the difficulty of obtaining a **larger number (around 40 or 50 images per category compared with 600 samples per category in the natural landscape and object image dataset ImageNet) of images for clinically rare skin diseases, we can only ensure a minimum of 20 images per category in Derm104, so this might be a limitation whereby trained neural network generalization is not good enough for certain disease categories**. We hope that those in the field of dermatology are able to collect and collate more images of skin diseases of clearer quality, especially in rare disease categories, with the development of automated health care.

Conclusion

In this paper, we introduce CDD-Net, a versatile plug-in module for clinical skin disease FSL classification, which performs well in different FSL approaches, including pretrained fine-tuning–based, meta-learning–based, and metric learning–based methods. CDD-Net presents a context feature–fusion module and dual-attention module to extract discriminating texture features and emphasize contributive region and channels. We also present Derm104, a new clinical

skin disease dataset with significant coverage for rare diseases and a reliable annotation between primary species and subspecies, as our benchmark. CDD-Net performed better on metric learning–based methods and has an accuracy improvement of up to 9.14 percentage points. Our ablation study validated its flexible structure and the importance of each component. As a future direction, we will investigate how to locate and extract skin lesion features of different sizes and shapes on complex and various backgrounds and summarize a better extraction logic, specifically for clinical skin backgrounds.

Data Sharing

The experimental dataset combines data from two sources. The first part was obtained from the public dermatology atlas website www.med126.com/pf. The second part is from the available dataset SD-198, collected by Sun et al.[23]

Ethics and Consent

Experimental data were selected from SD-198 and www.med126.com/pf. These two datasets are public and freely available for researchers, They were assembled in accordance with the 1964 Declaration of Helsinki and later amendments or comparable ethical standards. The patients whose records are in the database have given approval for their use. Users can download relevant data free for research and publish relevant articles. Our study is based on open-source retrospective data that had already been collected by other researchers. No new samples were collected from any patients, so there are no ethical issues or conflicts of interest.

SD-198 is a public skin disease dataset provided by Sun et al[23] that contains 6783 images of various skin lesions. This dataset has already led to improvements in dermatology research and is publicly available for anyone in the research community to use. The website www.med126.com/pf is a medical forum for students and researchers for study and discussion. Clinical skin disease images in many classes are demonstrated for study or research and are also freely available to download. Our research is free from ethical issues, based on the "Approaches to ethical review of life sciences and medical research" in China, as shown in the following paragraph.

According to items 1 and 2 of Article 32 of the Measures for Ethical Review of Life Science and Medical Research Involving Human Subjects, which was reviewed by the National Science and Technology Ethics Committee, approved by the State Council of China, and jointly promulgated by the National Health Commission, Ministry of Education, Ministry of Science and Technology and the State Administration of Traditional Chinese Medicine on February 18, 2023, stipulates the following: 1. The research is conducted using legally obtained public data or data generated through observation and does not interfere with public behavior. 2. Use of anonymized data for research. For more information on China's guidelines on research ethics and governance, please see https://www.gov.cn/zhengce/zhengceku/2023-02/28/content_5743658.htm.

Funding

Disclosure

The authors declare no conflict of interest.

References

1. Hay RJ, Johns NE, Williams HC, et al. The global burden of skin disease in 2010: an analysis of the prevalence and impact of skin conditions. *J Invest Dermatol*. 2014;134(6):1527–1534. doi:10.1038/jid.2013.446
2. Gonzalez-Castro V, Debayle J, Wazaefi Y, et al. Automatic classification of skin lesions using color mathematical morphology-based texture descriptors. Paper presented at: Twelfth International Conference on Quality Control by Artificial Vision; 2015.
3. Badano A, Revie C, Casertano A, et al. Consistency and standardization of color in medical imaging: a consensus report. *J Dig Imag*. 2015;28:41–52. doi:10.1007/s10278-014-9721-0
4. Jafari MH, Karimi N, Nasr-Esfahani E, et al. Skin lesion segmentation in clinical images using deep learning. Paper presented at: 2016 23rd International conference on pattern recognition (ICPR); 2016.
5. Yang J, Sun X, Liang J, Rosin PL. Clinical skin lesion diagnosis using representations inspired by dermatologist criteria. Paper presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018.

6. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. Paper presented at: International conference on machine learning; 2015.

7. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. Paper presented at: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015.

8. Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge. *Int J Comput Vis*. 2015;115:211–252. doi:10.1007/s11263-015-0816-y

9. Junayed MS, Islam MB, Jeny AA, Sadeghzadeh A, Biswas T, Shah AS. ScarNet: development and validation of a novel deep CNN model for acne scar classification with a new dataset. *IEEE Access*. 2021;10:1245–1258. doi:10.1109/ACCESS.2021.3138021

10. X-y Z, Wu X, F-f L, et al. The application of deep learning in the risk grading of skin tumors for patients using clinical images. *J Med Syst*. 2019;43:1–7.

11. Xie Y, Zhang J, Xia Y, Shen C. A mutual bootstrapping model for automated skin lesion segmentation and classification. *IEEE Transac Med Imag*. 2020;39(7):2482–2493. doi:10.1109/TMI.2020.2972964

12. Khan MA, Zhang Y-D, Sharif M, Akram T. Pixels to classes: intelligent learning framework for multiclass skin lesion localization and classification. *Comput Electr Eng*. 2021;90:106956.

13. Balaji VR, Suganthi ST, Rajadevi R, Kumar VK, Balaji BS, Pandiyan S. Skin disease detection and segmentation using dynamic graph cut algorithm and classification through Naive Bayes classifier. *Measurement*. 2020;163:107922. doi:10.1016/j.measurement.2020.107922

14. Roy K, Chaudhuri SS, Ghosh S, Dutta SK, Chakraborty P, Sarkar R. Skin Disease detection based on different Segmentation Techniques. Paper presented at: 2019 international conference on opto-electronics and applied optics (Optronix); 2019.

15. Lake B, Salakhutdinov R, Gross J, Tenenbaum J. One shot learning of simple visual concepts. Paper presented at: Proceedings of the annual meeting of the cognitive science society; 2011.

16. Xiao J, Xu H, Fang D, Cheng C, Gao H. Boosting and rectifying few-shot learning prototype network for skin lesion classification based on the internet of medical things. *Wireless Networks*. 2021;2021:1–15.

17. Lee K, Cavalcanti TC, Kim S, et al. Multi-task and few-shot learning-based fully automatic deep learning platform for mobile diagnosis of skin diseases. *IEEE J Biomed Health Inform*. 2022;27(1):176–187. doi:10.1109/JBHI.2022.3193685

18. Liu X-J, K-l L, H-y L, W-h W, Z-y C. Few-shot learning for skin lesion image classification. *Multimedia Tools Appl*. 2022;81(4):4979–4990. doi:10.1007/s11042-021-11472-0

19. Zhu W, Liao H, Li W, Li W, Luo J. Alleviating the incompatibility between cross entropy loss and episode training for few-shot skin disease classification. Paper presented at: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference. Lima, Peru; 2020:232020.

20. Shuhan LI, Li X, Xu X, Cheng K-T. Sub-cluster-aware network for few-shot skin disease classification. *arXiv preprint arXiv*. 2022;2022:1.

21. Prabhu V, Kannan A, Ravuri M, Chaplain M, Sontag D, Amatriain X. Few-shot learning for dermatological disease diagnosis. Paper presented at: Machine Learning for Healthcare Conference; 2019.

22. Mahajan K, Sharma M, Vig L. Meta-dermdiagnosis: few-shot skin disease identification using meta-learning. Paper presented at: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops; 2020.

23. Sun X, Yang J, Sun M, Wang K. A benchmark for automatic visual classification of clinical skin disease images. Paper presented at: Computer Vision–ECCV 2016: 14th European Conference. Amsterdam, The Netherlands; 2016: 142016.

24. Ravi S, Larochelle H. Optimization as a model for few-shot learning. Paper presented at: International conference on learning representations; 2017.

25. Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning. *Advan Neural Inform Process Sys*. 2017;2017:30.

26. Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. Paper presented at: International conference on machine learning; 2017.

27. Raghu A, Raghu M, Bengio S, Vinyals O. Rapid learning or feature reuse? Towards understanding the effectiveness of maml. *arXiv preprint arXiv*. 2019;2019:1.

28. Finn C, Yu T, Zhang T, Abbeel P, Levine S. One-shot visual imitation learning via meta-learning. Paper presented at: Conference on robot learning; 2017.

29. Sun Q, Liu Y, Chua T-S, Schiele B. Meta-transfer learning for few-shot learning. Paper presented at: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2019.

30. Chen W-Y, Liu Y-C, Kira Z, Wang Y-CF, Huang J-B. A closer look at few-shot classification. *arXiv preprint arXiv*. 2019;2019:1.

31. Tian Y, Wang Y, Krishnan D, Tenenbaum JB, Isola P. Rethinking few-shot image classification: a good embedding is all you need? Paper presented at: Computer Vision–ECCV 2020: 16th European Conference. Glasgow, UK; 2020:162020.

32. Li W, Wang L, Xu J, Huo J, Gao Y, Luo J. Revisiting local descriptor based image-to-class measure for few-shot learning. Paper presented at: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2019.

33. Rodrguez P, Laradji I, Drouin A, Lacoste A. Embedding propagation: smoother manifold for few-shot classification. Paper presented at: Computer Vision–ECCV 2020: 16th European Conference. Glasgow, UK; 2020:162020.

34. Liu B, Cao Y, Lin Y, et al. Negative margin matters: understanding margin in few-shot classification. Paper presented at: Computer Vision–ECCV 2020: 16th European Conference. Glasgow, UK; 2020:162020.

35. Ballerini L, Fisher RB, Aldridge B, Rees J. A color and texture based hierarchical K-NN approach to the classification of non-melanoma skin lesions. *Color Med Image Anal*. 2013;2013:63–86.

36. Codella NCF, Gutman D, Celebi ME, 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). Paper presented at: 2018 IEEE 15th international symposium on biomedical imaging; 2018.

37. Yi X, Walia E, Babyn P. Unsupervised and semi-supervised learning with categorical generative adversarial networks assisted by Wasserstein distance for dermoscopy image classification. *arXiv preprint arXiv*. 2018;2018:1.

38. Han SS, Kim MS, Lim W, Park GH, Park I, Chang SE. Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *J Invest Dermatol*. 2018;138(7):1529–1538. doi:10.1016/j.jid.2018.01.028

39. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Paper presented at: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015.

40. Manzoor K, Majeed F, Siddique A, et al. A lightweight approach for skin lesion detection through optimal features fusion. *Comput Mater Continua*. 2022;70(1):1617–1630. doi:10.32604/cmc.2022.018621

41. Ma S, Tang J, Guo F. Multi-task deep supervision on attention R2U-net for brain tumor segmentation. *Front Oncol*. 2021;11:704850. doi:10.3389/fonc.2021.704850

42. Wei Z, Shi F, Song H, Ji W, Han G. Attentive boundary aware network for multi-scale skin lesion segmentation with adversarial training. *Multimedia Tools Appl*. 2020;79:27115–27136. doi:10.1007/s11042-020-09334-2

43. Kasmi R, Mokrani K. Classification of malignant melanoma and benign skin lesions: implementation of automatic ABCD rule. *IET Image Process*. 2016;10(6):448–455. doi:10.1049/iet-ipr.2015.0385

44. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: visual explanations from deep networks via gradient-based localization. Paper presented at: Proceedings of the IEEE international conference on computer vision; 2017.