SM Dermatology Journal ISSN: 2575-7792

Original Article

GCProNet: Graph-Based Continual Prototypical Networks for Few-Shot Classification of Rare Skin Diseases

Abdulrahman Noman*, Zou Beiji*, Chengzhang Zhu, Mohammed Alhabib, Adnan Saeed and Raeed Al-Sabri

Central South University, China

Abstract

Prototypical networks have emerged as an effective approach for few-shot learn- ing, particularly in classifying rare skin diseases, by leveraging deep neural networks to create a feature space for classifying new, unseen classes. Nev- ertheless, obstacles such as the inability to fully retain or leverage long-term knowledge and inaccurate prototype estimation due to limited data reduce its ability to generalize well to new or unseen classes. To overcome these chal- lenges, we introduce GCProNet: Graph-Based Continual Prototypical Networks, a novel framework aimed at improving few-shot classification of rare skin dis- eases. GCProNet integrates support samples into the prototype network and captures the relational structure between these samples. The model transfers knowledge across tasks by adopting a continual learning strategy to improve classification accuracy. Initially, Convolutional Neural Networks extract fea- ture representations, which are then enhanced by graph-based techniques to capture dependencies among support samples. These graph-aggregated features are preserved through Gated Recurrent Units (GRUs) to facilitate continuous task learning. The resulting expanded feature space, enriched by both previous task knowledge and relational dependencies, is subsequently utilized within the Prototypical Network to generate more precise class prototypes, particularly for challenging new and rare classes. Experimental results demonstrate that GCProNet surpasses existing models, achieving 80.5% accuracy on the ISIC 2018 dataset, 86.12% on the Derm7pt dataset, and 92.63% on the SD-198 dataset under 5-shot learning conditions. These results demonstrate the strong potential of GCProNet in enhancing skin disease classification when data availability is limited.

INTRODUCTION

Significant progress in deep learning has greatly impacted the healthcare sector, particularly in medical image analysis. These techniques have demonstrated substantial promise in automating the diagnostic process from medical images, enhancing the precision and speed of multiple diagnostic tasks. Despite this, dermatology continues to pose challenges for deep learning models, primarily due to the vast diversity of skin conditions, especially rare ones, and the restricted availability of annotated datasets for less prevalent diseases, as illustrated in Figure 1. Furthermore, the manual annotation of medical images is a labor-intensive task prone to inaccuracies, making it challenging to produce the large annotated datasets necessary to train effective models for skin disease classification. These obstacles limit the successful application of deep learning in dermatology, particularly in few-shot learning scenarios where only a small number of labeled samples are available. Few-shot learning (FSL) has emerged as a viable solution to overcome the problem of scarce labeled data. Metric-based FSL models, such as Prototypical Networks (ProtoNets) [4], have shown considerable success in medical diagnostics by learning from a small number of labeled samples and generalizing effectively. In dermatology, the Meta-DD [5], model applied meta-learning to tackle the issue of limited data and class imbalance, while the MetaDerm [6], model addressed the difficulty of diagnosing rare diseases with limited annotated images. The SCAN algorithm [7], further improved FSL for skin disease classification by capturing intra-class variations through sub-clustered representations. While these methods have made strides, they still face substantial challenges related to data scarcity, inaccurate prototype estimations, and bias towards overrepresented classes, which restrict their capacity to classify a wide variety of skin diseases effectively. To resolve these limitations, we introduce GCProNet, a novel hybrid Prototypical Network integrating Graph Neural Networks (GNNs) for relational feature modeling and Gated Recurrent Units (GRUs) for continual knowledge adaptation. By modeling feature vectors as graph nodes and encoding their interrelations as edges, GCProNet effectively cap- tures rich contextual dependencies to enhance prototype representations. The incorporation of GRUs mitigates catastrophic forgetting, facilitating robust knowledge retention across tasks and improving generalization to novel skin disease categories. Extensive evaluation on benchmark datasets—SD- 198, Derm7pt, and ISIC 2018—demonstrates that GCProNet surpasses state-of-the-art methods, establishing it as a scalable and effective framework for few-shot dermatological classification. Our core contributions are:

1. We introduce GCProNet, an advanced architecture that reformulates the prototypical network to resolve critical few-shot dermatological classification challenges—data scarcity and cross- task catastrophic forgetting—through synergistic integration of graph relational modeling and recurrent knowledge preservation.

2. GCProNet integrates CNNs, GNNs, and GRUs to model contextual dependencies and retain task knowledge, to enrich the feature space, and to provide a more robust class representation.

3. GCProNet achieves state-of-the-art performance on dermatological benchmarks (ISIC 2018 [1], Derm7pt [2], SD-198 [3]), with significant accuracy gains demonstrating efficacy for few-shot diagnostic applications.

Submitted: 24 November 2025 | **Accepted:** 12 December 2025 | **Published:** 13 December 2025

***Corresponding author:** Abdulrahman Noman, Central South University, China

Zou Beiji, Central South University, China

Citation: Noman A, Beiji Z, Zhu C, Alhabib A, Saeed A, Al-Sabri R (2025) GCProNet: Graph-Based Continual Prototypical Networks for Few-Shot Classification of Rare Skin Diseases. SM Dermatol J 9: 10.

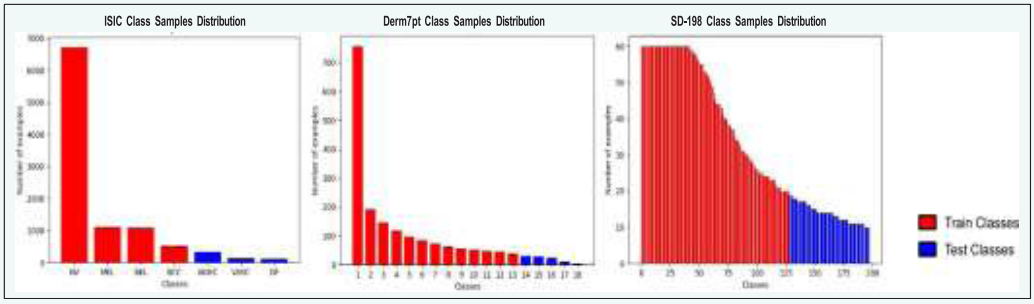

Figure 1: Skin disease datasets ISIC 2018 [1], Derm7pt [2], and SD-198[3] exhibit long-tail distributions. Models train on head classes (common diseases, red) and test on scarce tail classes (rare diseases, blue), creating a significant few-shot learning challenges.

RELATED WORKS

Few-shot Learning FSL

Few-shot learning (FSL) is a powerful approach for classifying new categories with limited labeled examples, leveraging knowledge from base classes to generalize to novel ones. FSL methods are broadly categorized into meta-gradient learning, metric-based approaches, and graph neural networks (GNNs), each offering unique advantages for training with limited data. *Meta-gradient Learning:* This approach optimizes model parameters through gradient-based methods tailored to specific tasks. Methods like Meta-LSTM [8], use Long Short-Term Memory networks to adapt model parameters for efficient learning, while MAML [9], learns optimal initial parameters through Stochastic Gradient Descent (SGD), applicable across supervised and reinforcement learning. Reptile [10], simplifies MAML using first-order gradients, and SNAIL [11], integrates temporal convolutions and attention mechanisms for dynamic task adaptation. *Metric-based Methods:* These methods classify based on feature vectors and distance metrics. Siamese networks [12], measure similarity between samples, while Prototypical Networks [1], generate class prototypes by averaging feature vectors and classify based on distance to these prototypes. The Relation Network [13], learns optimal similarity metrics for query-support comparisons. *Graph Neural Networks* (GNNs): GNNs capture relationships in graph- structured data, enhancing FSL performance. Early work [14] used GNNs to classify query samples by updating node information in a graph structure. TPN [15], introduced a transductive method using top-k graphs for label propagation from support to query sets. Later work [16] improved these models by integrating prototype-based label propagation. SPNP [17] and EGNN [18] incorporated structural graph information to tackle challenges like catastrophic forgetting and generalization, while Bayesian GNNs [19] enabled continual learning and task adaptation. Building on these methods, we propose GCProNet combines Graph Neural Networks and Gated Recurrent Units to enhance prototype generation and preserve task knowledge, achieving robust few-shot classification for skin diseases. This hybrid approach leverages relational reasoning and knowledge retention to improve dermatological classification performance.

FSL for Skin Disease Classification

Few-shot learning (FSL) for skin disease classification leverages transfer learning and meta-learning techniques to improve performance with limited data. Meta-learning approaches, such as gradient- based [20,21], and metric-based methods [5,6], have shown strong results. In transfer learning, Dai et al. [22], introduced a dual-encoder architecture that integrates large and small-scale datasets for better feature extraction in medical tasks. Xiao et al. [23], proposed a multitask framework with contrastive learning to enhance prototype networks for skin disease classification. Prabhu et al. [24], improved Prototypical Networks with a clustering approach for multiple prototypes per disease class, though it requires predefined sub-clusters. Shuhan Li et al. [7], introduced dynamic sub-clustering for more flexible feature encoding, especially for rare conditions. Self-supervised learning techniques [25,26], further enhance FSL models by utilizing unlabeled data. In gradient-based meta-learning, Li et al. [20], improved MAML with Difficulty-Aware Meta-Learning (DAML) to adapt to task complexity, while Singh et al. [21], enhanced the Reptile model with regularization techniques for better performance. For metric-based methods, Mahajan et al. [5], integrated group-equivariant convolutions into Prototypical Networks, enabling better feature invariance for dermatology. Desingu Kar et al. [6], proposed a meta-training technique for dermoscopic images, improving feature embedding for rare diseases.

GCProNet enhances Prototypical Networks by integrating Graph Neural Networks (GNNs) and memory units, addressing the challenges of few-shot skin disease classification. By capturing relational dependencies between support samples and incorporating continual learning, GCProNet improves prototype accuracy and generalization, especially for rare dermatological conditions.

METHODOLOGY

Problem Definition

Severe data scarcity—particularly for rare dermatological pathologies—poses fundamental challenges for classification, necessitating models that generalize to novel categories under minimal supervision. Formally, we define mutually exclusive datasets: a training dataset D_{train} for model learning and a testing dataset D_{test} containing novel classes ($D_{\text{train}} \cap D_{\text{test}} = \emptyset$). Within the N-way K-shot framework, a support set $S = \{(x_i, y_i)\}_{i=1}^{N \times K}$ provides limited supervision, while a query set $Q = \{(x_i, y_i)\}_{i=1}^{M}$ (unseen during training) evaluates generalization. The core challenge is leveraging $S \subset D_{\text{train}}$ to accurately classify novel dermatological categories in D_{test} given only K exemplars per class. The GCProNet model addresses this by expanding the feature space with graph-based methods that capture relational dependencies and enhancing feature extraction through CNNs. Additionally, GCProNet employs continual learning to preserve and transfer knowledge across tasks, improving its ability to classify skin diseases from limited training data.

The Model

GCProNet integrates graph-convolutional feature enhancement with prototypical networks to advance few-shot dermatological classification. As shown in Figure 2 and 3, our architecture overcomes traditional limitations through relational context modeling and flexible knowledge transfer.

Feature Extraction: GCProNet employs a CNN embedding backbone

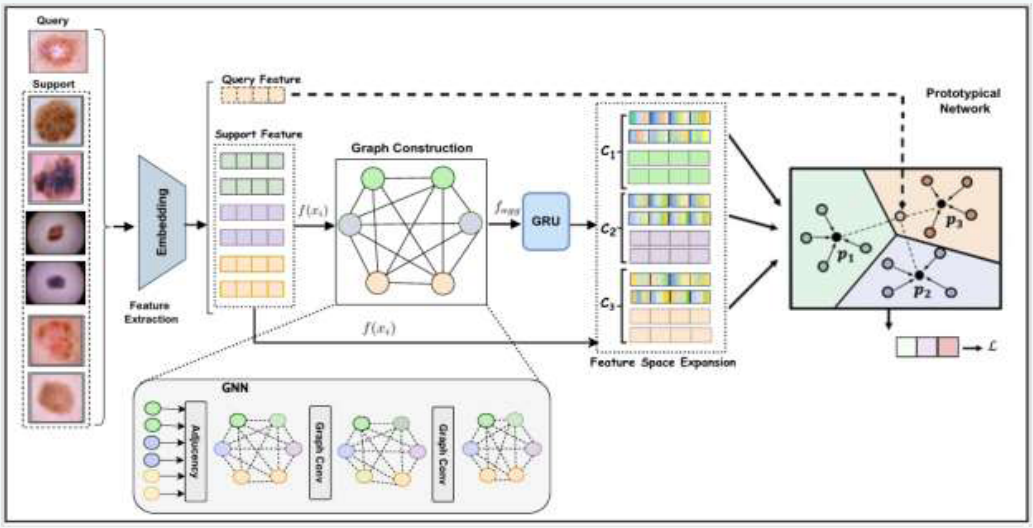

Figure 2: GCProNet architecture for few-shot dermatological diagnosis. In 3-way 2-shot tasks, a CNN backbone extracts dermoscopic features, dynamically aggregated via graph-structured propaga- tion (edge-learned relationships) and fused with GRU-stabilized embeddings. Optimized prototypes P_1, P_2, P_3 enable metric-based classification, addressing data scarcity through graph-guided knowl- edge transfer

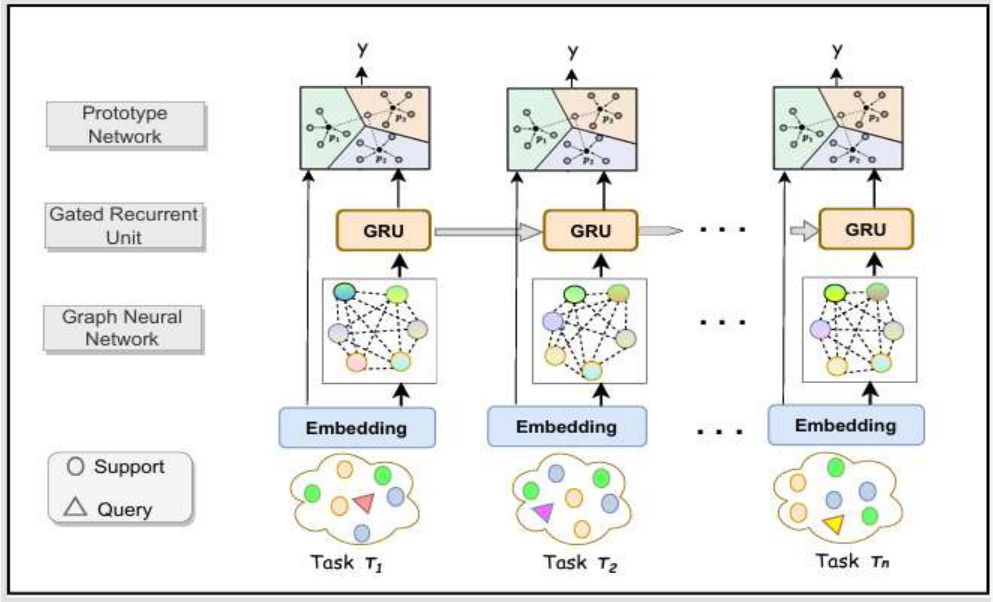

Figure 3: GCProNet integrates prototypical networks with GCN-based feature propagation and GRU-gated knowledge preservation for continual dermatological classification in sequential tasks *(T_1-T_n)*.

[27], for discriminative feature extraction, utilizing architectures such as Conv4, Conv6, or Wide ResNet (WRN) to extract discriminative features from dermatological support and query images, formally defined in Equation 1:

$$f(x_i) = \mathrm{CNN}(x_i; \theta) \tag{1}$$

In this context, $f(x_i)$ is the feature vector for the input image xi, while θ denotes the learnable parameters of the CNN. The architecture typically consists of convolutional layers to capture spatial information, blending layers to reduce dimensionality and improve generalization, and fully connected layers that combine the extracted features into the final representation $f(x_i)$. This feature vector effectively encodes the critical visual attributes of skin diseases, such as texture, shape, and pattern, which are essential for subsequent classification tasks. In particular, the quality and precision of the extracted features have a direct bearing on the success of subsequent graph-based feature enhancement and the classification process within the Prototypical Network framework. Consequently, GCProNet's overall performance relies heavily on CNN's ability to extract high-quality, discriminative features that accurately capture the disease characteristics.

Graph Construction and Continual Knowledge Transfer: Following the extraction of features using CNN, the next step involves constructing a graph from the representations of the characteristics of the support set.

This begins with the formation of an adjacency matrix *A*, which enables the integration of characteristics and the relationships between models. This aggregation enhances the feature vectors, strengthening their representational capacity and robustness for downstream tasks.

Graph Construction: Support image feature vectors, represented as $X = \{x_1, x_2, \ldots, x_n$ serve as graph nodes, with each node corresponding to $f(x_i)$ The adjacency matrix A encodes the pairwise relationships between these nodes, where each element A_{ij} measures the similarity between the nodes *i* and *j*. The specific formulation of this similarity is provided in Equation 2:

$$A_{ij} = \text{Similarity}(f(x_i), f(x_j)) = \exp(-\frac{d(f(x_i), f(x_j))}{2\sigma^2}) \quad (2)$$

where d(·,·) denotes a distance metric, typically the Euclidean distance, employed to evaluate similarity between feature vectors $(f(x_i), f(x_j))$ via a Gaussian kernel. The parameter σ^2 serves as a scaling factor, consistent with methodologies such as embedding propagation [28], and TPN [15].

Feature Aggregation: The model applies the feature aggregation function $f_{agg}(x_i)$ to each node after the graph has been established. This process is implemented through the utilization of graph based neural network techniques, specifically graph-convolutional networks (GCNs) [29]. Equation 3 mathematically describes the aggregation of characteristics:

$$f_{agg}^{l+1}(x_i) = \sigma\left(\hat{D}^{-1/2}\hat{A}\hat{D}^{-1/2}f(x_i)^l W\right) \quad (3)$$

where $\hat{A} = A + I_N$ denotes the adjacency matrix with self-loops, ˆD is its corresponding degree matrix, W the learnable weight matrix, and σ a non-linear activation (e.g., ReLU). After two propagation layers, node embeddings incorporate multi-hop neighborhood context, synthesizing localized patterns and global structural relationships into discriminative representations.

Continual Knowledge Transfer via GRU Memory Units: After aggregating graph-based features, GCProNet employs Gated Recurrent Units (GRUs) [30], to enable effective knowledge transfer across tasks, addressing the issue of catastrophic forgetting in sequential few-shot dermatological classification, as illustrated in Figure 3. GRUs accomplish this by employing two adaptive gating mechanisms: 1. *Update Gate (z_t)*: Controls the retention of the previous hidden state h_{t-1}, thereby preserving long-term dependencies that are critical for maintaining inter-task feature consistency. 2. *Reset Gate (r_t)*: Manages the integration of historical context into the current candidate state $\tilde{h}_t$, enabling the model to discard obsolete patterns and incorporate novel lesion classes effectively. The dynamics of the GRU are governed by the following set of equations:

$$\begin{aligned} z_t &= \sigma(W_z[h_{t-1}; f_{\text{agg}}(x_i)]), \\ r_t &= \sigma(W_r[h_{t-1}; f_{\text{agg}}(x_i)]), \\ \tilde{h}_t &= \tanh(W_h[r_t \odot h_{t-1}; f_{\text{agg}}(x_i)]), \\ h_t &= (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t, \end{aligned} \quad (4)$$

where [;] denotes vector concatenation, ⊙ represents element-wise multiplication, and $f_{agg}(x_i) \in R^d$ refers to the graph-refined feature representation from task *t*. The learnable parameters W_z, W_r, and W_h are meta-optimized to balance the contributions of historical knowledge h_{t-1} and incoming task embeddings $f_{agg}(x_i)$, and σ denotes the sigmoid activation function. By iteratively refining h_t, the GRUs construct a dynamically evolving feature space that retains critical dermatological patterns (such as lesion texture and border irregularity) across tasks. This process effectively prevents feature degradation when incrementally learning rare classes with fewer than or equal to five samples. The final GRU-enhanced embeddings h_t are subsequently propagated to the prototypical network, where they stabilize the prototype computation by ensuring both inter-class separability and intra-class compactness—key characteristics for robust few-shot classification.

Feature Space Expansion: GCProNet's architecture systematically expands the feature space of the support set, facilitating richer and more discriminative representations essential for accurate skin disease classification. This enhanced representation results from the fusion of CNN-extracted features, refined through graph-based methods and sequentially encoded using Gated Recurrent Units (GRUs), mathematically formulated as follows:

$$f_{exp}(x_i) = f(x_i) + h_t \quad (5)$$

Here, $f_{exp}(x_i)$ denotes the expanded feature vector for each image x_i integrating the initial CNN features $f(x_i)$ with the feature enhancements provided by the GRU h_t. Unlike conventional prototype networks that rely on simple feature averaging, GCProNet integrates individual features with their contextual interrelations within the support set, substantially enhancing classification accuracy. This approach is especially advantageous in few-shot learning, where limited data necessitates leveraging both local and global feature contexts to maximize performance and generalizability.

Prototypical Network and Loss Optimization: In the GCProNet framework, the Prototypical Network [4-31], utilizes the enriched feature space to generate class prototypes, each defined as the mean vector of expanded support set features for class *c*. These prototypes serve as representative centroids crucial for accurate classification, as follows:

$$P_c = \frac{1}{N_c}\sum_{i=1}^{N_c} f_{exp}(x_i) \quad (6)$$

where N_c denotes the number of samples in class *c*, and $f_{exp}(x_i)$ represents the enriched feature vectors of i_{th} the sample. For the query images, feature vectors are extracted via the CNN, ensuring distinct representations for comparison. These query features are then evaluated against the class prototypes within the expanded feature space, employing the Euclidean distance to quantify similarity. The distance between the feature vector of query image q_j and prototype P_c is defined as:

$$d_{jc} = \|f(q_j) - P_c\| \quad (7)$$

This distance metric forms the basis of the classification decision, where the class assigned to each query corresponds to the nearest prototype. To express this decision probabilistically, a softmax function is applied over the negative distances, yielding a probability distribution over all classes:

$$\text{Prob}(y = c \mid q_j) = \frac{e^{-d_{jc}}}{\sum_{k=1}^{C} e^{-d_{jk}}} \quad (8)$$

Model training is guided by the cross-entropy loss, computed as:

$$\mathcal{L}(q_j) = -\sum_{c=1}^{C} \mathbb{1}(y_j = c) \log \text{Prob}(y = c \mid q_j) \quad (9)$$

Figure 4: Sample images illustrating skin lesions from diverse datasets: (a) ISIC 2018 dataset, (b) Derm7pt dataset (clinical and dermoscopic images), and (c) SD-198 dataset.

where $\mathbb{1}(\cdot)$ denotes the indicator function, equal to 1 when the true label of q_i is c, and 0 otherwise. By employing this enhanced methodology within the Prototypical Network, underpinned by an improved feature space, GCProNet significantly enhances its ability to accurately classify skin diseases, especially in few-shot learning scenarios.

EXPERIMENTS

To evaluate GCProNet, we conducted experiments on two widely recognized benchmark datasets: ISIC-2018 [1] and Derm7pt [2]. These collections, which comprise a diverse array of dermatological images, serve as critical resources for assessing and enhancing the model's capability in skin lesion classification.

Datasets

ISIC-2018 dataset: The ISIC-18 [1], comprises 10,015 dermoscopic images of skin lesions, classified into seven categories. For training, four classes (9,431 images) were used, while three classes (584 images) served as the test set. To ensure computational efficiency, all images were resized from 600×450 to 224×224 pixels. For few-shot learning experiments, four classes were designated for meta-training, and three classes were used for meta-testing, capturing the complexity inherent in clinical skin lesion diagnosis. Representative samples from the dataset are illustrated in Figure 4(a).

Derm7pt Dataset: The Derm7pt dataset [2] contains over 2,000 clinical and dermoscopic images across 20 lesion categories. After resizing from 768×512 to 224×224 pixels, standard augmentations (cropping, rotation, and flipping) were applied. Excluding the "miscellaneous" and "melanoma" classes due to insufficient samples, 18 categories were divided into 13 training and 5 testing classes. The dataset was further split into a base set (1,892 images, 40–698 per class) and a novel set (114 images, 10–34 per class), reflecting real-world data scarcity challenges. as illustrated in Figure 4(b).

SD-198 Dataset: The SD-198 dataset [3] comprises 6,584 clinical images spanning 198 dermato logical conditions. Images were resized from 1640 × 1130 to 224 × 224 pixels for experimental consistency. For evaluation, we tested 70 rare-disease classes (≤ 20 images/class), while 20 classes (60 images/class) were used for training, ensuring a fair assessment under limited-data conditions. Representative sample images from the dataset are illustrated in Figure 4(c).

Implementation Details

Network Architecture: The proposed framework combines three embedding architectures: Conv4, Conv6, and Wide ResNet (WRN-28-10) [32]. Conv4 and Conv6 are shallow CNNs with 4 and 6 convolutional layers, respectively, each utilizing 64 kernels (3×3), ReLU activation, batch normalization [33], and 2×2 max-pooling. The deeper WRN-28-10 features 28-layer residual blocks with a widening factor of 10, improving gradient flow and feature representation. A graph neural network (GNN) with two graph convolutional layers processes adjacency matrices to aggregate node-level features, capturing relational patterns vital for few-shot classification. Gated Recurrent Units (GRUs) [30], consolidate cross-task knowledge, enhancing model generalization across diverse few-shot tasks.

Experimental Settings: The model was implemented in PyTorch and trained on an NVIDIA A100- SXM4-40GB GPU. We followed standard few-shot learning protocols for the ISIC-2018, Derm7pt, and SD-198 datasets. For ISIC-2018, we used 2-way 1/3/5-shot configurations, while Derm7pt and SD-198 employed 2-way 1/5-shot setups. Each episode sampled 5 query images, and performance was assessed using query-set accuracy and macro-F1 scores. Training followed the N-way K-shot episodic paradigm, with the Adam optimizer [43], a learning rate of 10^{-4}, and weight decay of 10^{-3}. The models were trained for 1000 epochs on ISIC-2018, 1500 epochs on Derm7pt, and 2000 epochs on SD-198 to ensure sufficient adaptation to each dataset's challenges.

Experimental Analysis

We assess GCProNet on **ISIC-2018** [1], **Derm7pt** [2], and **SD-198** [3] under 1/3/5-shot settings. It couples a GNN for relational reasoning over the support set with a GRU-based memory for continual retention, mitigating prototype bias and catastrophic forgetting.

Result on the ISIC-18 dataset, as shown in Table 1, GCProNet achieved 80.5% accuracy and 83.2% AUC in the 5-shot configuration, significantly surpassing baseline models like MetaDerm [6], (73.1% accuracy) and PCN [24], (73.0% accuracy). In the 1-shot configuration, GCProNet demonstrated robust generalization with 67.3% accuracy, outperforming Meta-DD [5], (59.3%) and MetaDerm (58.9%). This superior performance is attributed to the model's GNN-driven relational aggregation and GRU-powered continual knowledge transfer, which mitigate common issues in rare dermatological datasets [7-24].

On the **Derm7pt dataset**, which includes clinically annotated dermoscopic images as shown in Table 2, GCProNet excelled at inferring diagnostic patterns, such as pigment networks and vascular structures. Using the WRN-28-10 backbone, GCProNet achieved 86.12% accuracy in the 5-shot configuration, outperforming SCAN [7] (82.57%) and MetaDerm (67.0%). In the 1-shot configuration, GCProNet reached 68.19% accuracy, surpassing SCAN (62.80%) [7]. The hybrid architecture of GCProNet, which integrates CNNs for localized feature extraction, GNNs for relational aggregation, and GRUs for continual knowledge retention, enabled it to significantly outperform baseline models. Notably, GCProNet improved F1-scores by 2.2–4.3% over MetaDerm and Multi-task [34], demonstrating better precision-recall trade-offs, which are crucial in medical diagnostics.

For the **SD-198 dataset** as shown in Table 3, which spans 198 skin diseases, GCProNet achieved a 5-shot accuracy of 92.63% and 1-shot

Table 1: Average accuracy and area under the curve (AUC) for 2-way 1/3/5-shot on **ISIC-18** with a Conv6 backbone.

Model	Backbone	Metric		1-shot	3-shot	5-shot
Meta-DD [2]		Avg. Avg.	Accuracy AUC	59.3% 61.6%	67.9% 70.2%	73.0% 75.7%
MetaDerm [3] Multi-task [32]	Conv6	Avg. Avg. Avg. Avg.	 Accuracy AUC Accuracy AUC 	58.9% 69.1% 62.7% —	68.3% 72.3% 73.1% —	73.1% 80.6% 77.6% —
MTDD [33]		Avg. Avg.	 Accuracy AUC	60.5% —	69.8% —	76.3% —
GCProNet	Conv6	Avg. Accuracy Avg. AUC		**67.3%** **69.6%**	**75.8%** **79.6%**	**80.5%** **83.2%**

Table 2: Results on Derm7pt: GCProNet vs. baselines under 2-way 1-shot and 2-way 5-shot; reported as F1 and accuracy

Method	Backbone	2-way Accuracy	1-shot F1-score	2-way Accuracy	5-shot F1-score
PCN [24]	Conv4	59.98%	58.54%	70.62%	71.85%
SCAN [7]		61.42%	61.90%	72.58%	74.05%
GCProNet		**63.13%**	**63.08%**	**74.25%**	**74.90%**
Meta-DD [5]	Conv6	61.8%	—	76.9%	—
MetaDerm [6]		62.3%	—	67.0%	—
SCAN [7]		62.80%	63.75%	76.65%	73.60%
CDD-Net [35]		59.46%	—	73.78%	—
Multi-task [33]		62.70%	—	77.83%	—
MTDD [34]		63.35%	63.98%	75.61%	76.11%
GCProNet		**68.19%**	**67.36%**	**77.65%**	**77.02%**
NCA [36]	WRN-28-10	56.32%	56.41%	67.18%	68.13%
Baseline [37]		59.43%	59.61%	74.28%	75.26%
S2M2-R [38]		61.37%	61.52%	79.83%	80.69%
NegMargin [39]		58.00%	57.50%	70.12%	71.07%
PT+NCM [40]		60.92%	61.12%	74.33%	74.96%
BEM E-NCM [41]		60.40%	60.57%	72.63%	73.01%
EASY [42]		61.02%	61.25%	75.98%	76.43%
SCAN [7]		66.75%	67.71%	82.57%	83.73%
GCProNet		**70.08%**	**68.77%**	**86.12%**	**85.96%**

Table 3: Results on **SD-198**: GCProNet vs. baselines under **2-way 1-shot** and **2-way 5-shot**; reported as F1 and accuracy.

Method	Backbone	2-way Accuracy	1-shot F1-score	2-way Accuracy	5-shot F1-score
PCN [24]	Conv4	70.03%	70.78%	84.95%	85.87%
SCAN [7]		77.12%	78.00%	90.22%	91.01%
GCProNet		**77.89%**	77.51%	86.05%	85.82%
Meta-DD [5]		65.3%	-	83.7%	-
SCAN [7]	Conv6	76.75%	77.64%	87.45%	88.28%
GCProNet		**77.30%**	**77.84%**	**87.60%**	87.38%
NCA [37]		71.27%	71.27%	83.30%	84.23%
Baseline [38]		75.72%	76.64%	88.95%	84.23%
S2M2-R [39]		76.42%	77.51%	90.32%	90.97%
NegMargin [40]		76.85%	77.98%	89.92%	90.65%
PT+NCM [41]	WRN-28-10	78.25%	78.86%	90.33%	90.90%
BEM E-NCM [42]		78.32%	78.70%	90.48%	90.94%
EASY [43]		78.80%	79.44%	90.87%	91.43%
SCAN [7]		80.20%	81.21%	91.48%	92.08%
GCProNet		**80.23%**	79.98%	**92.63%**	**92.48%**

accuracy of 80.23% using the WRN-28-10 backbone, surpassing SCAN (91.48%) and EASY [41] (90.87%). The GRU component played a critical role in reducing forgetting rates, ensuring that GCProNet retained discriminative features across sequential tasks. However, when using the Conv4 backbone, GCProNet underperformed compared to SCAN in the 5-shot setting (86.05% vs. 90.22%), suggesting that lightweight backbones may not fully leverage relational dependencies without further optimization. This highlights the need for task-specific adaptations in graph-based learning to handle more complex feature spaces [40-42]. Finally, GCProNet represents a significant advancement in few-shot learning for dermatology, combining graph-based relational learning with continual knowledge retention. It addresses key challenges in rare disease classification, including sparse data, intra-class variability, and evolving diagnostic criteria [20,21]. As shown in Figure 5, the experimental results consistently highlight GCProNet's superior performance compared to baseline methods across all evaluated datasets.

This demonstrates GCProNet's robust few-shot generalization, enabling clinical deployment in resource-limited settings with scarce data.

Ablation studies

Component Analysis: The study highlights the critical impact of individual GCProNet components on dermatopathological classification. Removing the graph-based module (Prototype w/o Graph) led to performance drops across all datasets, particularly in the 1-shot case (ISIC-2018 accuracy reduced from 67.3% to 62.8%). Similarly, eliminating prototype learning (Graph w/o Prototype) further decreased accuracy, underlining the importance of relational dependencies for effective generalization. The exclusion of Gated Recurrent Units (GRUs) (GProNet w/o GRU) resulted in a notable performance decline, especially in the 5-shot scenario (ISIC-2018 accuracy fell from 80.5% to 77.46%), demonstrating GRUs' role in preserving and transferring knowledge to prevent catastrophic forgetting. In contrast, integrating all components into the full GCProNet model significantly improved results: 80.5% accuracy on ISIC-2018 (5-shot), 77.65% on Derm7pt, and 87.60% on SD-198. These findings underscore the synergy of graph-based relational learning, prototype-based classification, and GRUs in addressing the challenges of sparse data, intra-class variability, and long-term knowledge retention.

Grad-CAM for Enhanced Feature Interpretation in GCProNet: Gradient-weighted Class Activation Mapping (Grad-CAM) is a powerful method for visualizing the regions within an image that are most influential in a model's decision-making process. In the case of GCProNet, Grad-CAM generates heatmaps that are overlaid on the input images, visually

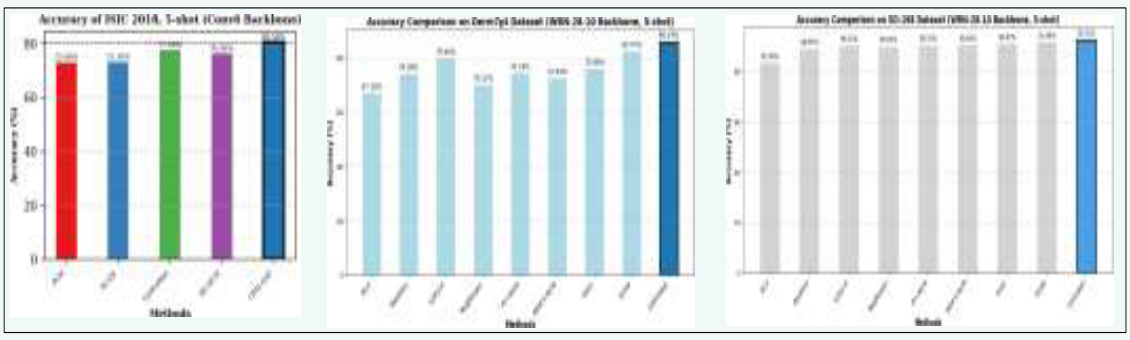

Figure 5: Comparative evaluation of GCProNet against baseline methods on three dermatology datasets (ISIC 2018, Derm7pt, and SD-158) under a 5-shot learning setting..

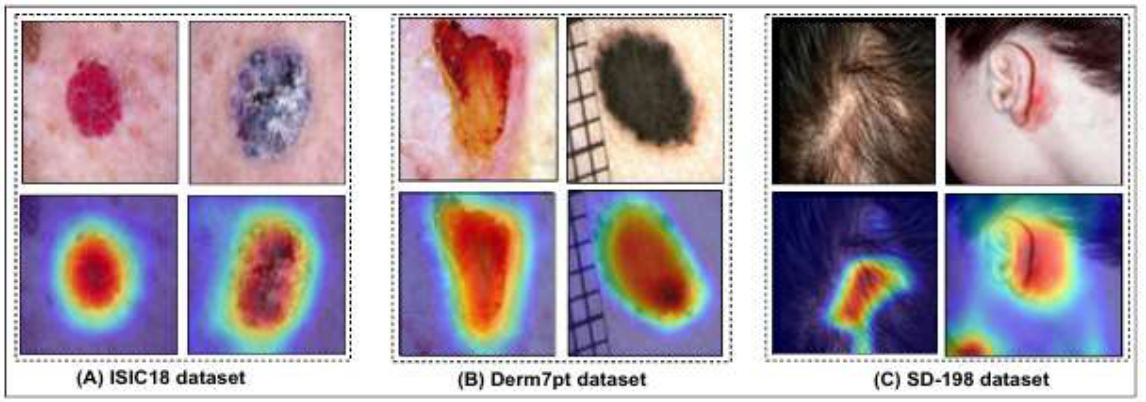

Figure 6: Grad-CAM visualization applied to GCProNet for skin disease classification. The heatmaps highlight the most relevant regions of the input images, demonstrating the areas of the image that contribute most significantly to the model's classification decision.

Table 4: Ablation study comparing GCProNet's effectiveness in **2-way** few-shot learning tasks, employing a **Conv6** backbone across the ISIC-2018, Derm7pt, and SD-198 datasets

Method	ISIC-2018		Derm7pt		SD-198	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Prototype w/o Graph	62.8%	74.1%	62.5%	66.9%	69.7%	74.50%
Graph w/o Prototype	55.3%	73.6%	58.2%	64.6%	67.2%	71.78%
GProNet w/o GRU	63.02%	77.46%	64.43%	71.83%	73.25%	82.56%
Integrated (GCProNet)	67.3%	80.5%	68.19%	77.65%	77.30%	87.60.%

representing the areas that contribute most to the classification outcome. This technique provides enhanced transparency, which is essential in the medical field, where the interpretability of automated systems is critical. By identifying the key features that drive model predictions, Grad-CAM strengthens the reliability and trustworthiness of GCProNet, enabling clinical validation and offering evidence-based insights into the model's performance. As illustrated in Figure 6, this visualization fosters greater understanding and confidence in the automated classification of skin diseases.

Limitations and Future work: Despite strong few-shot performance in dermatological image classification, GCProNet remains sensitive to extreme class imbalance and pronounced intra-class variability. Future work will improve robustness to severe imbalance, advance relational modeling, and incorporate domain knowledge (e.g., lesion context and multi-modal cues) to further enhance generalization.

CONCLUSION

GCProNet overcomes ProtoNets' limitations through graph-relational reasoning and continual adaptation for data-scarce dermatological classification. By integrating a graph-based framework to capture relational dependencies among support samples and incorporating Gated Recurrent Units (GRUs) for knowledge retention, GCProNet significantly enhances both the accuracy and stability of dermatological diagnosis, particularly for rare skin conditions. Extensive evaluations on the Derm7pt, SD-198, and ISIC 2018 datasets demonstrate substantial improvements in classification performance, highlighting GCProNet's ability to overcome challenges such as data scarcity and prototype inaccura- cies. GCProNet sets a new benchmark for few-shot learning in medical applications, paving the way for more reliable diagnostic tools in dermatology and beyond.

ACKNOWLEDGEMENTS

This work is supported by the National Key R&D Program of China (2018AAA0102100), the Key Research and Development Program of Hunan Province (2022SK2054); the National Natural Science Foundation of China (No. 62376287); the International Science and Technology Innovation Joint Base of Machine Vision and Medical Image Processing in Hunan Province (2021CB1013); and the Natural Science Foundation of Hunan Province (No. 2022JJ30762).

REFERENCES

1. Codella N, Rotemberg V, Tschandl P, Celebi ME, Dusza S, Gutman D, et al. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic).arXiv preprint arXiv:1902.03368, 2019.
2. Kawahara J, Daneshvar S, Argenziano G, Hamarneh G. Seven-point checklist and skin lesion classification using multitask multimodal neural nets. IEEE J Biomed Health Informatics. 2018; 23: 538–546.
3. Sun X, Yang J, Sun M, Wang K. A benchmark for automatic visual classification of clinical skin disease images. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VI 14. Springer. 2016; 206–222.
4. Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning. Advances in neural information processing systems. 2017; 30.
5. Mahajan K, Sharma M, Vig L. 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.
6. Desingu K, Chandrabose PA. Few-shot classification of skin lesions from dermoscopic images by meta-learning representative embeddings. arXiv preprint arXiv:2210.16954. 2022.
7. Li S, Li X, Xu X, Cheng KW. Dynamic subcluster-aware network for few-shot skin disease classification. IEEE Transactions on Neural Networks and Learning Systems, 2023.
8. Ravi S, Larochelle H. Optimization as a model for few-shot learning. In International conference on learning representations. 2016.
9. Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adap- tation of deep networks. In International conference on machine learning. PMLR. 2017; 1126-1135.
10. Nichol A, Achiam J, Schulman J. On first-order meta-learning algorithms.arXiv preprint arXiv:1803.02999. 2018.
11. Mishra N, Rohaninejad M, Chen X, Abbeel P. A simple neural attentive meta-learner. arXiv preprint arXiv:1707.03141, 2017.
12. Koch G, Zemel R, Salakhutdinov R. Siamese neural networks for one-shot image recognition. In ICML deep learning workshop. Lille, 2015; 37.
13. Sung F, Yang Y, Zhang L, Xiang T, Torr PH, Hospedales TM. Learning to compare: Relation network for few-shot learning. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2018; 1199-1208.
14. Satorras VG, Estrach JB. Few-shot learning with graph neural networks. In International conference on learning representations, 2018.
15. Liu Y, Lee J, Park M, Kim S, Yang E, Hwang SJ, et al. Learning to propagate labels: Transductive propagation network for few-shot learning. In 7th International Conference on Learning Representations, 2019; 5.
16. Zhu H, Koniusz P. Transductive few-shot learning with prototype-based label propagation by iterative graph refinement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023; 23996-24006.
17. Lin X, Li Z, Zhang P, Liu L, Zhou C, Wang B, et al. Structure-aware prototypical neural process for few-shot graph classification. IEEE Transactions on Neural Networks and Learning Systems. 2022; 35: 4607:4621.
18. Kim J, Kim T, Kim S, Yoo CD. Edge-labeling graph neural network for few-shot learning. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019; 11–20.
19. Luo Y, Huang Z, Zhang Z, Wang Z, Baktashmotlagh M, Yang Y. Learning from the past: continual meta-learning with bayesian graph neural networks. In Proceedings of the AAAI Conference on Artificial Intelligence. 2020; 34: 5021–5028.
20. Wang L, Dou Q, Fletcher PT, Speidel S, Li S. Medical image computing and computer assisted intervention—miccai 2022. In Proceedings of the 24th International Conference, Strasbourg, France. 2021; 12901: 109–119.
21. Singh R, Bharti V, Purohit V, Kumar A, Singh AK, Singh SK. Metamed: Few-shot medical image classification using gradient-based meta-learning. Pattern Recognition. 2021; 120: 108111.
22. Dai Z, Yi J, Yan L, Xu Q, Hu L, Zhang Q, et al. Pfemed: Few-shot medical image classification using prior guided feature enhancement. Pattern Recognition. 2023; 134: 109108.
23. Xiao J, Xu H, Fang D, Cheng C, Gao HH. Boosting and rectifying few-shot learning prototype network for skin lesion classification based on the internet of medical things. Wireless Networks. 2023; 29: 1507–1521.
24. Prabhu V, Kannan A, Ravuri M, Chaplain M, Sontag D, Amatriain X. Few-shot learning for dermatological disease diagnosis. In Machine Learning for Healthcare Conference. PMLR. 2019; 532-552.
25. Chen D, Chen Y, Li Y, Mao F, He Y, Xue H. Self-supervised learning for few-shot image classification. In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE. 2021; 1745–1749.
26. Medina C, Devos A, Grossglauser M. Self-supervised prototypical transfer learning for few-shot classification. arXiv preprint arXiv:2006.11325. 2020.
27. Albawi S, Mohammed TA, Al-Zawi S. Understanding of a convolutional neural network. In 2017 international conference on engineering and technology (ICET). Ieee. 2017; 1-6.
28. Rodríguez P, Laradji I, Drouin A, Lacoste A. Embedding propagation: Smoother manifold for few-shot classification. In Computer Vision–ECCV 2020: 16th European Conference. Springer. 2020; 121–138.
29. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907, 2016.
30. Chung J, Gulcehre C, Cho KH, Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555, 2014.
31. Zhou F, Wang P, Zhang L, Wei W, Zhang Y. Revisiting prototypical network for cross domain few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023; 20061-20070.

32. Zagoruyko S, Komodakis N. Wide residual networks. arXiv preprint arXiv:1605.07146, 2016.

33. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International conference on machine learning. 2015; 37: 448–456.

34. Khurshid M, Vatsa M, Singh R. Multi-task explainable skin lesion classification. arXiv preprint arXiv:2310.07209, 2023.

35. Özdemir Z, Keles HY, Tanrıöver OO. Meta-transfer derm-diagnosis: Exploring few-shot learning and transfer learning for skin disease classification in long-tail distribution. arXiv preprint arXiv:2404.16814, 2024.

36. Chen T, Liu Q, Yang J. Few-shot classification with multiscale feature fusion for clinical skin disease diagnosis. Clinical, Cosmetic and Investigational Dermatology. 2024; 17: 1007–1026.

37. Wu Z, Efros AA, Yu SX. Improving generalization via scalable neighborhood component analysis. In Proceedings of the european conference on computer vision (ECCV). 2018; 685-701.

38. Chen WY, Liu YC, Kira Z, Wang Y-CF, Huang JB. A closer look at few-shot classification. arXiv preprint arXiv:1904.04232. 2020.

39. Mangla P, Kumari N, Sinha S, Singh S, Krishnamurthy B, Balasubramanian VN. 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.

40. Liu B, Cao Y, Lin Y, Li Q, Zhang Z, Long M, et al. Negative margin matters: Understanding margin in few-shot classification. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, Proceedings, Part IV 16. Springer. 2020; 438–455.

41. Hu Y, Gripon V, Pateux S. Leveraging the feature distribution in transfer-based few-shot learning. In International Conference on Artificial Neural Networks. Springer. 2021. 487-499.

42. Hu Y, Pateux S, Gripon V. Squeezing backbone feature distributions to the max for efficient few-shot learning. MDPI, 2022; 15: 147.

43. Bendou Y, Hu Y, Lafargue R, Lioi G, Pasdeloup B, Pateux, et al. Easy—ensemble augmented-shot-y-shaped learning: State-of-the-art few-shot classification with simple components. J Imaging. 2022; 8: 179.

44. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.