Improving Feature Representations for Few-shot Classification of Neglected Tropical Diseases on Dark Skin Tones

Felix Obete[0009−0000−9234−8618], Ian Raymond Osolo[0000−0001−9891−6161], and Terhemba Michael-Ahile[0009−0000−8030−7309]

Uganda Christian University, Mukono, Uganda. info@ucu.ac.ug
www.ucu.ac.ug

Abstract. Transfer learning addresses limited labeled data in deep learning by leveraging pre-trained model knowledge. In extreme data scarcity scenarios, such as rare medical conditions, few-shot classification becomes essential, relying critically on feature representation quality. A key challenge is producing discriminative, transferable representations when target domains diverge significantly from source domains. This study investigates strategies to improve pre-trained model representations for few-shot classification of skin-related Neglected Tropical Diseases (NTDs) on dark skin tones (Fitzpatrick Scale IV–VI). We propose a sequential framework combining supervised domain adaptation to align representations with darker skin dermatology images, followed by supervised contrastive learning (SupCon) to enhance inter-class separability under data scarcity. Using DenseNet-161 and ViT-B/16 as feature extractors, we evaluated performance over 100 randomly sampled support/query episodes per setting and reported mean accuracy with 95% confidence intervals and paired significance tests. Domain adaptation significantly improved the pre-trained baseline for ViT-B/16 across all five few-shot settings (Wilcoxon, $p < 0.001$) and for DenseNet-161 in three of five settings ($p < 0.05$). The highest mean accuracy, 81.1% (95% CI [78.9, 83.3]), was achieved by ViT-B/16 in the 3-way 5-shot setting after supervised contrastive learning. Although its additional gains beyond domain adaptation were generally modest, supervised contrastive learning complemented the pipeline by refining learned feature representations and achieving the highest mean performance in a select few-shot settings. These findings demonstrate the effectiveness of combining domain adaptation with supervised contrastive learning to improve few-shot NTD recognition on dark skin tones.

Keywords: Neglected Tropical Skin Diseases · Few-Shot Classification · Feature Representations · Transfer Learning · Domain Adaptation.

1 Introduction

Recent advances in machine learning and deep learning have impacted several domains, including dermatological diagnosis [1]. Skin diseases are the most common cause of human illness, affecting about 900 million people at any time, and

rank as the fourth leading cause of non-fatal illness; conditions such as skin cancer can be fatal without early diagnosis [2–4]. Others, such as Neglected Tropical Diseases (NTDs) of the skin, are prevalent in impoverished tropical communities facing acute shortages of dermatologists, and if left unattended can cause disfigurement, chronic disability, and morbidity [5–7].

Despite their success, deep learning models perform substantially worse on underrepresented dark skin tones [8–10], a challenge that is especially severe for Neglected Tropical Skin Diseases [11]. While transfer learning has been applied to dermatology diagnosis [12–14], its use for NTDs on dark skin tones may face three challenges: (i) the domain gap between general-purpose source datasets (e.g., ImageNet) and dermatology-specific target domains limits feature transferability; (ii) intra-class appearance shifts across skin tones, where the same condition appears markedly differently on lighter versus darker skin [15]; and (iii) the scarcity or absence of many NTD categories in existing dermatology datasets, resulting in classes unseen during conventional supervised training.

Prior work on NTD diagnosis includes [11], who benchmarked CNNs, transformer based, and medical foundation models on a multi-modal NTD dataset, reaching a best accuracy of 61.68% with PanDerm. To improve classification accuracy without access to large data, we refine feature representations from general-purpose pre-trained models through domain adaptation [16] and supervised contrastive learning [17], and implement few-shot classification [18] to address sample scarcity, rather than the conventional classification used in [5, 11]. We first evaluate which pre-trained models are the most suitable dermatology feature extractors, following [19]. We then apply tone-aware domain adaptation, via partial fine-tuning of select layers [20, 21], focused on Fitzpatrick skin tones IV-VI, followed by supervised contrastive learning to enforce distinct feature geometry. Few-shot evaluation of the resulting representations follows the premise that a good embedding is all that is needed [22], rather than relying on complex episodic few-shot architectures.

Our main contribution is improved few-shot classification accuracy for Neglected Tropical Skin Diseases on dark skin tones without episodic pre-training, which typically requires large amounts of data. To the best of our knowledge, this is the first application of partial fine-tuning combined with supervised contrastive learning for this task under extreme data scarcity.

2 Related Work

2.1 Feature Extractor Evaluation

Selecting an appropriate backbone remains a key consideration in transfer learning. Several dermatology studies [5, 8, 12] use pre-trained models as feature extractors, freezing backbone parameters and fine-tuning only a classifier head, a setup not designed to fully characterize the extractor itself, but sufficient for comparing which pre-trained model best suits a given dataset [8, 11]. Study by [19] confirms that pre-trained models perform differently across domains, and

[23] benchmarked a broad suite of pre-trained models and tasks to ease this choice.

2.2 Domain Adaptation

Domain adaptation in dermatology has been approached in several ways. In their work, [24] fine-tuned on the small DDI dataset augmented with the larger HAM10000 skin cancer dataset, risking catastrophic forgetting of pre-trained weights while HAM10000's cancer-focused images leave out thousands of other conditions. [25] proposed target-guided dynamic mixup (TGDM), pairing a Mixup-3T classification network with a dynamic ratio generation network (DRGN) to synthesize intermediate domain data, but assumes constant availability of both source and target data. [26] combined transfer and meta-learning, adapting a backbone via scaling, translation, and challenging positive/negative support samples; performance degrades as shot count increases, and the approach assumes uniform target-domain tasks, an assumption that breaks down in dermatology, where the same condition appears markedly differently across the Fitzpatrick Skin Tone (FST) scale from type I to VI. [20] instead approximated target batch-normalization statistics as a linear combination of source and support statistics (LCCS), but this hypothesis does not hold when transferring ImageNet weights to dermatology, particularly dark skin tones, given the large source-target domain gap. [21] reviewed fine-tuning strategies for vision transformers, showing that updating only Feed-Forward network parameters achieves near-best results on ViT-B/16, an approach closely aligned with the one used in our study.

2.3 Supervised Contrastive Learning

Supervised contrastive learning (SupCon), introduced in [17], pulls together embeddings of same-class samples while pushing apart different classes, achieving state-of-the-art results on large-scale classification. Its two-stage pipeline—contrastive pre-training of an encoder followed by cross-entropy fine-tuning of a linear classifier on frozen representations decouples representation learning from classification, enhancing transferability. In dermatology, [27] applied SupCon to CNNs for skin lesion classification, mitigating class imbalance, while [28] introduced a multi-modal "Contrast-based Consistent Representation Disentanglement" technique that preserves uniform lesion attributes across modalities without requiring more data. Evaluation of SupCon's effect on feature geometry draws on [29, 30], using cosine-based metrics: mean pairwise cosine distance/similarity and mean nearest centroid cosine distance/similarity.

2.4 Few-shot Classification

A study by [31] showed that transfer learning outperforms sophisticated meta-learning for few-shot classification. In dermatology, [12, 13] extend this to data-scarce, long-tailed distributions by training on an initial class set before adapting

to tail classes, focusing on long-tail imbalance rather than intra-class domain shift.

3 Dataset

The Fitzpatrick17k dataset [8] contains 16,577 clinical images from two dermatology atlases (12,672 DermaAmin, 3,905 Atlas Dermatologico) across 114 disease classes, annotated by Fitzpatrick skin type (FST I–VI). The dataset is imbalanced: light tones (FST I–III) account for 69% of samples versus 31% for darker tones (FST IV–VI). The SD-198 dataset [32] comprises 6,584 clinical images across 198 skin disease classes, acquired under diverse imaging conditions but predominantly representing light skin tones, with limited dark-skin examples.

We assembled our evaluation dataset from two sources for adequate dark-skin representation across all classes. Scabies (18 images) came from Fitzpatrick17k's held-out, dark-tone subset; its "ulcers" class was excluded for being predominantly lighter-skinned. The remaining seven classes: Buruli ulcer (14), yaws (16), mycetoma (19), lymphatic filariasis (10), post-kala-azar dermal leishmaniasis (17), leprosy (15), and onchocerciasis (15), were sourced by drawing roughly four images per class from the WHO NTD training guide [4], with the remainder from publicly available online medical and dermatological resources. Labels were independently reviewed and confirmed by two collaborating physicians, following the same public-resource compilation approach as Fitzpatrick17k [8]. In total, the evaluation set comprises 124 images across 8 classes, all visually assessed as Fitzpatrick IV–VI. These 124 images were strictly held out for few-shot evaluation only, excluded from feature-extractor evaluation, domain adaptation, and SupCon training. The remaining Fitzpatrick17k IV–VI images were used for these stages and split into separate training and validation sets.

4 Methodology

We sequentially evaluated pre-trained feature extractors, domain adaptation (DA), and supervised contrastive learning (SupCon) for neglected tropical disease classification on dark skin tones. Few-shot evaluation after each stage quantified gains from representation refinement.

4.1 Few-shot Classification

For each of the random episodes, the dataset was partitioned into support (D_S) and query (D_Q) sets using 1-5 labeled samples per class. Following prototype-based classification [12, 22, 34], class prototypes, c_s were computed [12, 22, 34] as

$$c_s = \frac{1}{|D_S^s|} \sum_{x_i \in D_S^s} f_\phi(x_i). \tag{1}$$

In (1), $f_\phi(x_i)$ is the feature vector function. Given a query sample, $x_q \in D_q$. The goal was to obtain the support class where the normalized mean feature vector c_s is most similar to the normalized feature vector of the image, x_q. Obtained by selecting a class whose prototype feature vector has the highest cosine similarity with the normalized query feature vector, given in (2) below.

$$\hat{y}_q = \arg\max \left(\frac{f_\phi(x_q) \cdot c_s}{\|f_\phi(x_q)\| \, \|c_s\|} \right) \tag{2}$$

4.2 Feature Improvement Pipeline

Domain adaptation. During backpropagation, partial fine-tuning updated only trainable parameters θ_{train}, while frozen parameters θ_{frozen} remained fixed, for a cross entropy loss L and learning rate α:

$$\theta_i^{t+1} = \theta_i^t - \alpha M_i \nabla L(\theta^t), \tag{3}$$

In (3) $M_i \in 0, 1$ denotes whether parameter θ_i is trainable, ∇ is gradient loss function [21, 36].

Supervised contrastive learning. Domain-adapted DenseNet-161 and ViT-B/16 models were further refined using SupCon loss [17] to improve intra-class compactness and inter-class separation:

$$L = \sum_{i=1}^{N} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp((z_i \cdot z_p)/\tau)}{\sum_{j \neq i} \exp((z_i \cdot z_j)/\tau)}. \tag{4}$$

Here,L is SupCon Loss, z denotes L_2-normalized embeddings, $P(i)$ is the set of all indices in the batch sharing the same label. $|P(i)|$ is the count of positive samples for the anchor. $j \neq i$ is all other samples in the batch and τ is the temperature parameter [17, 27].

5 Experiments, Results and Discussions

5.1 Experiment Setup

Identical preprocessing, augmentation, and training protocols were maintained across feature extraction (FE), domain adaptation (DA), and supervised contrastive learning (SupCon). Few-shot evaluation used 8-way 5-shot, 5-way 5-shot, 3-way 5-shot, 3-way 2-shot, and 3-way 1-shot settings, each assessed over 100 randomly sampled support/query episodes drawn without replacement from a pooled set of 124 images across 8 NTD classes, with mean accuracy and 95% confidence intervals reported per setting; paired Wilcoxon signed-rank tests [37] compared consecutive refinement stages (FE - DA, DA - SupCon) on matched episode draws. For FE, images were resized, center-cropped, and normalized; frozen backbones with linear classification heads were trained with Adam ($\alpha =$ 0.001, batch size =32, patience =5) and selected by weighted F1. For DA,

DenseNet-161 updated Block 4, batch normalization, and the classifier, while ViT-B/16 updated only feed-forward layers, preserving attention and normalization; adapted models initialized SupCon training. SupCon used an L2-normalized projection head: DenseNet-161 with $\tau = 0.1$, $\alpha 0.00005$, $128 - dprojection$; ViT-B/16 with $\tau = 0.07$, $\alpha 0.00003$. Projection heads were discarded post-training, retaining representations for downstream few-shot evaluation.

5.2 Results

Feature extractor evaluation on Fitzpatrick17k and SD-198 showed dataset-specific performance differences. Overall, ViT-B/16 achieved the highest mean weighted F1 score among transformer models of 53.94%, while DenseNet-161 was the best CNN baseline at 44.57%, as shown in Table 1

Table 1. Average performance across datasets

Model	Fitzpatrick17k Weighted F1	SD-198 Weighted F1	Mean
ViT-B/16	0.5642	0.5147	**0.53945**
ViT-B/32	0.5312	0.4753	0.50325
DenseNet-161	0.4446	0.4467	**0.44565**
VGG-16	0.4475	0.3472	0.39735
EfficientNet-B0	0.3900	0.4219	0.40595

Table 2 summarizes domain adaptation results. Fine-tuning DenseNet-161 Block 4 and the classifier improved weighted F1 from 0.3985 to 0.5545, while partial adaptation of ViT-B/16 by updating only FFN layers achieved the best performance overall weighted F1 of 0.5838.

SupCon training on domain-adapted DenseNet-161 and ViT-B/16 converged rapidly, with early stopping at epochs 16 and 14 respectively. As shown in Table 3, DenseNet-161's mean pairwise and nearest-centroid cosine distances marginally increased (0.0146 and 0.0137), with cosine similarity decreasing correspondingly. ViT-B/16 instead showed a substantial decrease in both distances (0.206131 and 0.172622).

Few-shot results across the three refinement stages, evaluated over 100 randomly sampled episodes per setting with 95% confidence intervals, are summarized

Table 2. Domain adaptation of DenseNet-161 and ViT-B/16 on Fitzpatrick17k dataset (scale IV to VI)

Model	Architecture Section	Parameters	Val Accuracy	Weighted F1	Macro F1
DenseNet-161	Head	236,363	42.50%	0.3985	0.3445
DenseNet-161	Block4 + Head	9,723,083	56.56%	0.5545	0.4708
ViT-B/16	Head	82,283	42.50%	0.3952	0.3180
ViT-B/16	FFN + Head	56,751,467	58.66%	0.5838	0.4795

Table 3. Feature space geometry analysis after supervised contrastive learning (SupCon)

Model	Metrics	Pre	Post	Change
DenseNet-161	Mean Pairwise Cos Dist	0.3493	0.3639	0.0146
DenseNet-161	Mean Nearest Centroid Cos Dist	0.2154	0.2291	0.0137
DenseNet-161	Mean Pairwise Cos Similarity	0.6507	0.6361	-0.0146
DenseNet-161	Mean Nearest Centroid Cos Similarity	0.7846	0.7709	-0.0137
ViT-B/16	Mean Pairwise Cos Dist	0.970664	0.764532	-0.206131
ViT-B/16	Mean Nearest Centroid Cos Dist	0.434588	0.261967	-0.172622
ViT-B/16	Mean Pairwise Cos Similarity	0.029336	0.235468	0.206131
ViT-B/16	Mean Nearest Centroid Cos Similarity	0.565412	0.738033	0.172622

Table 4. Results of few-shot evaluation across the three stages of the study. Values report mean accuracy ± 95% confidence interval over $n = 100$ randomly sampled episodes per setting.

Model(Stage)	8W 5S	5W 5S	3W 5S	3W 2S	3W 1S
DenseNet-161(Pre-trained)	55.2 ± 1.4	61.3 ± 1.8	71.7 ± 2.0	62.3 ± 2.8	51.9 ± 3.1
DenseNet-161(DA)	57.0 ± 1.4	63.0 ± 1.9	73.7 ± 2.2	63.1 ± 2.7	55.8 ± 3.1
DenseNet-161(SupCon)	57.3 ± 1.4	63.8 ± 2.0	74.4 ± 2.2	64.6 ± 2.4	55.6 ± 3.1
ViT-B/16(Pre-trained)	53.2 ± 1.5	62.2 ± 2.1	73.9 ± 2.5	64.5 ± 2.7	56.3 ± 3.1
ViT-B/16(DA)	64.9 ± 1.3	71.7 ± 1.8	80.1 ± 2.1	70.1 ± 2.7	62.9 ± 3.2
ViT-B/16(SupCon)	66.1 ± 1.3	70.3 ± 1.5	81.1 ± 2.2	69.1 ± 2.6	62.5 ± 3.0

in Table 4. Domain adaptation produced a statistically significant improvement over the pre-trained baseline for ViT-B/16 in all five settings (Wilcoxon, $p < 0.001$), and for DenseNet-161 in three of five settings ($p < 0.05$), with the remaining two showing a positive but non-significant trend. Supervised contrastive learning's further improvement over domain adaptation was smaller and reached significance in only one setting per architecture (DenseNet-161: 3-way 2-shot, $p = 0.037$;ViT-B/16: 5-way 5-shot, $p = 0.019$, where the shift was a modest decrease rather than a gain). The best overall result across both architectures and all stages, 81.1% (95% CI [78.9, 83.3]), was achieved by ViT-B/16 after SupCon in 3-way 5-shot.

5.3 Discussions

Feature extractor evaluation confirmed that DenseNet-161 and ViT-B/16 consistently outperformed other models within their respective families, with ViT-B/16 the overall best performer, reflecting both the dense connectivity and hierarchical feature reuse suited to fine-grained skin texture, and the broader strength of transformer-based models in dermatology.

Domain adaptation via partial fine-tuning Table 2 improved both models despite the constrained fine-tuning set of 4,965 images across 107 classes. Episodic evaluation with paired significance testing (Section 5.2) confirmed this improvement was statistically reliable for ViT-B/16 across every setting tested, and for

DenseNet-161 in three of five settings, indicating domain adaptation is the more consistently impactful refinement stage for both architectures, with a stronger and more uniform effect for the transformer. Supervised contrastive learning substantially reorganized the feature geometry for ViT-B/16 more than that of DenseNet-161. Table 3 shows that for ViT-B/16, mean pairwise cosine distance and mean nearest-centroid cosine distance dropped by 0.2061 and 0.1726, respectively, compared to the minimal increase (0.0146 and 0.0137) in the metrics for DenseNet-161. This divergence mirrors the architectures' differing reliance on large-scale pretraining for stable representations. This geometric reorganization, however, did not consistently translate into significant few-shot accuracy gains beyond domain adaptation: SupCon's mean shift over domain adaptation was typically under 1.5 percentage points and inconsistent in direction, reaching significance in only two of ten architecture–setting combinations (DenseNet-161 3-way 2-shot, $p = 0.037$; ViT-B/16 5-way 5-shot, a modest decrease, $p = 0.019$). This suggests embedding-space compactness and downstream few-shot performance, while related, are not tightly coupled at this sample size, and that domain adaptation not contrastive refinement is the dominant, statistically robust contributor to few-shot performance in this pipeline.

These findings align with [19] in confirming that pre-trained model effectiveness is dataset-dependent, extending this to pigmentation-sensitive dermatology. Our domain adaptation results are consistent with [21]'s partial fine-tuning approach, preserving learned lower-level features under extremely limited data. Our best result (81.1%, 95% CI [78.9, 83.3]) is substantially higher than both [24]'s MedViT benchmark (73% on a broader, non-NTD dataset) and the 61.68% reported by [11] for PanDerm on the comparable NTD task. We note these comparisons are illustrative rather than direct. Both [11] and [24] evaluated conventional supervised classification with different class counts and data splits, whereas our figure reflects prototype based few-shot accuracy under 3-way 5-shot protocol. The comparison suggests that domain-specific alignment yields substantial gains beyond pre-training alone. These results support [22]'s premise that a good embedding, rather than a complex episodic architecture, is sufficient for few-shot classification, extending this claim, unlike [22, 35], to a setting with substantial domain shift.

Limitations include the restricted set of architectures evaluated, the sequential rather than jointly optimized application of domain adaptation and contrastive learning, and modest class sample sizes. Because each class was represented by only 10–19 images, the space of distinct support/query partitions was bounded, and repeated 100-episode resampling draws overlapping rather than independent samples from this pool. The study focused solely on Fitzpatrick IV–VI tones; equitable AI requires continued evaluation across all skin types.

6 Conclusion and Future Work

This study shows that domain adaptation, applied to pre-trained models, substantially improves few-shot classification of Neglected Tropical Skin Diseases on

dark skin tones, with statistically significant gains confirmed via paired episodic evaluation for ViT-B/16 across all tested settings and for DenseNet-161 in the majority of settings. Supervised contrastive learning's further contribution beyond domain adaptation was smaller by comparison.

Accuracy of 81.1% (95% CI [78.9, 83.3]) for ViT-B/16 in 3-way 5-shot after SupCon, demonstrates that domain-specific alignment alone can drive substantial, data-efficient recognition gains for rare NTD classes unseen in conventional training. Future work could incorporate larger, progression-aware dermatology datasets, investigate joint rather than sequential optimization of domain adaptation and contrastive learning, and extend this framework to other domains with hierarchical domain shift and data scarcity.

Acknowledgments. We thank Dr. Martha Asiimwe (Hope and Healing Center, Kiwanyi) for confirming diagnostic labels for part of the evaluation dataset, and Dr. Apollo Mukwaya for facilitating image access.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. J. Zhang, F. Zhong, K. He, M. Ji, S. Li, and C. Li, "Recent advancements and perspectives in the diagnosis of skin diseases using machine learning and deep learning: A review," *Diagnostics*, vol. 13, no. 23, 2023.
2. K. A. Muhaba, K. Dese, T. M. Aga, F. T. Zewdu, and G. L. Simegn, "Automatic skin disease diagnosis using deep learning from clinical image and patient information," *Skin Health and Disease*, vol. 2, no. 1, 2022.
3. D. Seth, K. Cheldize, D. Brown, and E. E. Freeman, "Global burden of skin disease: inequities and innovations," *Current Dermatology Reports*, vol. 6, 2017.
4. WHO, *Recognizing neglected tropical diseases through changes on the skin: A training guide for front-line health workers*, 2018.
5. R. R. Yotsu, Z. Ding, J. Hamm, and R. E. Blanton, "Deep learning for AI-based diagnosis of skin-related neglected tropical diseases: A pilot study," *PLoS Neglected Tropical Diseases*, vol. 17, no. 8, 2023.
6. J. Olobo-Okao and P. Sagaki, "Leishmaniasis in Uganda: Historical account and a review of the literature," *Pan African Medical Journal*, vol. 18, 2014.
7. J. H. Kolaczinski et al., "Neglected tropical diseases in Uganda: The prospect and challenge of integrated control," *Trends in Parasitology*, vol. 23, pp. 485–493, 2007.
8. M. Groh, C. Harris, L. Soenksen, F. Lau, R. Han, A. Kim, A. Koochek, and O. Badri, "Evaluating deep neural networks trained on clinical images in dermatology with the Fitzpatrick 17k dataset," in *CVPR Workshops*, 2021, pp. 1820–1828.
9. L. Kamulegeya, J. Bwanika, M. Okello, D. Rusoke, F. Nassiwa, W. Lubega, D. Musinguzi, and A. Börve, "Using artificial intelligence on dermatology conditions in Uganda: a case for diversity in training data sets for machine learning," *African Health Sciences*, vol. 23, no. 2, pp. 753–763, 2023.
10. N. Munia and A.-A. Imran, "DermDiff: generative diffusion model for mitigating racial biases in dermatology diagnosis," arXiv preprint arXiv:2503.17536, 2025.

11. W. Janet, H. Xin, Z. Yunbei, A. Diabate, B. Vagamon, K. K. A. S. Sebastien, A. Y. Koffi, D. Zhengming, H. Jihun, and Y. Rie R., "eSkinHealth: A multimodal dataset for neglected tropical skin diseases," in *ACM MM*, 2025, pp. 12949–12956.
12. Z. Özdemir, H. Y. Keles, and Ö. Ö. Tanrıöver, "Meta-transfer derm-diagnosis: exploring few-shot learning and transfer learning for skin disease classification in long-tail distribution," 2024.
13. K. Mahajan, M. Sharma, and L. Vig, "Meta-DermDiagnosis: Few-shot skin disease identification using meta-learning," in *CVPR Workshops*, 2020, pp. 3142–3151.
14. S. Jain, U. Singhania, B. Tripathy, E. A. Nasr, M. K. Aboudaif, and A. K. Kamrani, "Deep learning-based transfer learning for classification of skin cancer," *Sensors*, vol. 21, no. 23, Art. no. 8142, 2021.
15. K. Fogelberg, S. Chamarthi, R. C. Maron, J. Niebling, and T. J. Brinker, "Domain shifts in dermoscopic skin cancer datasets: evaluation of essential limitations for clinical translation," *New Biotechnology*, vol. 76, pp. 106–117, 2023.
16. Y. Mansour, M. Mohri, and A. Rostamizadeh, "Learning and domain adaptation," in *ALT*, 2009.
17. P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, "Supervised contrastive learning," in *NeurIPS*, 2020.
18. W.-Y. Chen, Y.-C. Liu, Z. Kira, Y.-C. Wang, and J.-B. Huang, "A closer look at few-shot classification," in *ICLR*, 2019.
19. E. da Silva Puls, M. V. Todescato, and J. L. Carbonera, "An evaluation of pre-trained models for feature extraction in image classification," in *ICEIS*, 2024.
20. W. Zhang, L. Shen, W. Zhang, and C.-S. Foo, "Few-shot adaptation of pre-trained networks for domain shift," in *IJCAI*, 2022.
21. P. Ye, Y. Huang, C. Tu, M. Li, T. Chen, T. He, and W. Ouyang, "Partial fine-tuning: a successor to full fine-tuning for vision transformers," arXiv:2312.15681, 2023.
22. Y. Tian, Y. Wang, D. Krishnan, J. B. Tenenbaum, and P. Isola, "Rethinking few-shot image classification: a good embedding is all you need," in *ECCV*, 2020.
23. M. Goldblum, H. Souri, R. Ni, M. Shu, V. Prabhu, G. Somepalli, P. Chattopadhyay, M. Ibrahim, A. Bardes, J. Hoffman, R. Chellappa, A. G. Wilson, and T. Goldstein, "Battle of the backbones: a large-scale comparison of pretrained models across computer vision tasks," in *NeurIPS*, 2023.
24. S. A. Dip, K. H. I. Arif, U. A. Shuvo, I. A. Khan, and N. Meng, "Equitable skin disease prediction using transfer learning and domain adaptation," *AAAI Symposium Series*, 2024.
25. L. Zhuo, Y. Fu, J. Chen, Y. Cao, and Y.-G. Jiang, "TGDM: target guided dynamic mixup for cross-domain few-shot learning," in *ACM MM*, 2022.
26. R. Perera and S. Halgamuge, "Discriminative sample-guided and parameter-efficient feature space adaptation for cross-domain few-shot learning," in *CVPR*, 2024.
27. K. Chen, D. Zhuang, and J. Chang, "SuperCon: supervised contrastive learning for imbalanced skin lesion classification," arXiv:2202.05685, 2022.
28. Z. Wang, L. Zhang, X. Shu, Y. Wang, and Y. Feng, "Consistent representation via contrastive learning for skin lesion diagnosis," *Comput. Methods Programs Biomed.*, 2023.
29. Y. Wen, W. Liu, Y. Feng, B. Raj, R. Singh, A. Weller, M. Black, and B. Schölkopf, "Pairwise similarity learning is SimPLE," in *ICCV*, 2023.
30. A. Kutuzov and M. Giulianelli, "UiO-UvA at SemEval-2020 Task 1: contextualised embeddings for lexical semantic change detection," in *SemEval*, 2020, pp. 126–134.

31. A. Chowdhury, M. Jiang, S. Chaudhuri, and C. Jermaine, "Few-shot image classification: just use a library of pre-trained feature extractors and a simple classifier," in *ICCV*, 2021.
32. X. Sun, J. Yang, M. Sun, and K. Wang, "A benchmark for automatic visual classification of clinical skin disease images," in *ECCV*, 2016.
33. G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *CVPR*, 2017.
34. A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., "An image is worth 16×16 words: transformers for image recognition at scale," in *ICLR*, 2021.
35. G. S. Dhillon, P. Chaudhari, A. Ravichandran, and S. Soatto, "A baseline for few-shot image classification," arXiv:1909.02729, 2019.
36. Y. Akimoto, S. Shirakawa, N. Yoshinari, K. Uchida, S. Saito, and K. Nishida, "Adaptive stochastic natural gradient method for one-shot neural architecture search," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2019.
37. F. Wilcoxon, "Individual comparisons by ranking methods," Biometrics Bulletin, vol. 1, no. 6, pp. 80–83, 1945.