Artifact-aware Prompt Learning for Domain Generalization

Xibin Jia[1⋆], Wei Zhang[1], Shaowu Xu[1], and Chao Fan[1]

School of Computer Science, Beijing University of Technology, Beijing, China
jiaxibin@bjut.edu.cn

Abstract. Domain shifts induced by heterogeneous imaging artifacts remain a major obstacle to deploying dermoscopic image classification models in real-world clinical settings. Such artifacts introduce complex domain-specific variations that are difficult to address using domain-invariant representations alone. In this work, we propose an artifact-aware prompt-based framework for domain generalization in dermoscopic image classification, termed Artifact-aware Prompt Learning for Domain Generalization (APDG). During training, common imaging artifacts are explicitly modeled as multiple source domains, and domain-specific visual prompts are introduced within a Vision Transformer to encode heterogeneous artifact patterns. To mitigate domain bias while preserving useful domain knowledge, we further design a soft-mask-based feature disentanglement module that separates domain-invariant semantic features from domain-specific representations, together with a domain-specific prototype learning mechanism that structurally constrains disentangled features through joint domain supervision and auxiliary class guidance. At inference time, a test-time prompt fusion strategy is proposed to adaptively aggregate learned domain-specific knowledge without requiring target domain labels or additional training. Experiments on multiple unseen dermoscopic and clinical datasets demonstrate that the proposed method achieves improved robustness and generalization performance under severe artifact-induced domain shifts.

Keywords: Domain Generalization · Prompt Learning · Feature Disentanglement · Dermoscopic Image Classification.

1 Introduction

Dermoscopic image analysis plays an important role in melanoma diagnosis [8, 7]. Despite recent advances in deep learning, models trained on curated datasets often suffer from significant performance degradation in real-world clinical settings due to domain shifts between training and testing data [22, 2].

A major source of such shifts stems from heterogeneous imaging artifacts, including hairs, illumination variations, dark corners, and acquisition-related noise [1, 12]. These artifacts introduce domain-specific variations that are largely

⋆ corresponding author

irrelevant to lesion semantics but strongly influence feature distributions, causing models to overfit artifact-induced spurious correlations and generalize poorly across datasets.

Domain generalization (DG) seeks to learn models that generalize to unseen target domains without access to target data during training. Most existing DG approaches focus on enforcing domain-invariant representations by aligning features across source domains [24, 4, 16]. However, directly pursuing domain invariance in dermoscopic image classification faces two key challenges. First, skin lesions exhibit substantial variability in shape, color, and texture, and excessive alignment across heterogeneous domains may discard discriminative fine-grained information, leading to degraded performance. Second, since target-domain data are unavailable during training, domain-invariant features learned from source domains are not guaranteed to remain stable when encountering unseen artifact patterns at test time.

These limitations suggest that suppressing domain-specific information may be suboptimal. Instead, it is desirable to explicitly model artifact-induced domain variations while preserving lesion-relevant semantics, and to flexibly exploit learned domain knowledge under unknown target domains.

To this end, we propose a unified prompt-based framework for dermoscopic image domain generalization. During training, domain-specific visual prompts are introduced to encode heterogeneous artifact characteristics. A feature disentanglement mechanism is further employed to separate domain-specific and domain-invariant representations, while a prototype-guided learning strategy structures class–domain relationships and stabilizes the semantic space.

At inference time, we address the absence of domain labels by introducing a test-time prompt fusion strategy that adaptively aggregates domain-specific prompts based on prototype similarity, enabling robust generalization without target-domain supervision. Extensive cross-dataset experiments demonstrate that the proposed method consistently improves generalization performance on multiple unseen target domains.

The main contributions of this work are as follows: (1) We formulate dermoscopic image domain generalization from an artifact-aware perspective, explicitly modeling heterogeneous imaging artifacts as multiple source domains; (2) we propose a unified prompt-based framework integrating domain-specific prompts, feature disentanglement, and prototype-guided learning to preserve lesion semantics under artifact-induced shifts; (3) we introduce a test-time prompt fusion strategy that adaptively exploits learned domain knowledge without requiring target-domain labels; and (4) extensive experiments on multiple datasets validate the effectiveness of the proposed method under severe domain shifts.

2 Related Work

Domain Generalization in Dermoscopic Image Analysis. Domain generalization (DG) aims to learn models that generalize to unseen target domains without accessing target-domain data during training [19]. Traditional DG ap-

proaches include empirical risk minimization (ERM) [6], distributionally robust optimization (DRO) [17], and other robust optimization methods that have been systematically benchmarked in the DomainBed framework [6], as well as representation learning strategies such as SelfReg [11] and MMLD [13]. These methods have been widely adopted to improve robustness across different data distributions. However, in dermoscopic image analysis, DG is particularly challenging due to large variations in acquisition devices, imaging protocols, and artifact patterns across datasets [22]. A major yet under-explored challenge is artifact-induced domain shifts caused by heterogeneous imaging conditions (e.g., hairs, dark corners, gel bubbles) [1].

Prompt-based Domain Generalization. Prompt learning has recently emerged as an effective mechanism for adapting Transformer-based models to downstream tasks [9, 25]. By introducing learnable prompt tokens into the input sequence, prompt-based methods enable flexible task- or domain-specific adaptation while preserving the generalization capability of pretrained backbones. Several recent works have explored prompt-based strategies for domain generalization. DoPrompt introduces domain-specific prompts to encode domain characteristics [23], while EPVT leverages environment-aware prompt generation for improved cross-domain robustness [21]. More recently, PLDG further investigates prompt-driven learning by modeling latent domain variations through adaptive prompt generation [20]. While these methods demonstrate promising performance, they primarily rely on prompt adaptation mechanisms learned during training and do not explicitly enforce a structured disentanglement between domain-induced variations and lesion-relevant semantics, nor do they jointly leverage domain-invariant and domain-specific representations at inference time.

Test-time Adaptation. Test-time adaptation (TTA) focuses on improving model performance on unseen target domains by adapting model behavior during inference. A representative approach is Tent [18], which minimizes prediction entropy at test time without requiring target-domain labels. Although effective, many TTA methods involve iterative optimization and may introduce instability or additional computational overhead during deployment. Recently, prompt-based adaptation has been explored as a lightweight alternative for test-time robustness [5]. Instead of updating model parameters, prompt tuning or fusion strategies adjust prompt representations conditioned on test samples, enabling adaptive inference without explicit target-domain supervision. This perspective motivates the development of prototype-guided prompt fusion mechanisms that exploit learned domain knowledge while maintaining stable semantic representations.

3 Method

We study domain generalization for dermoscopic image classification. Let $\mathcal{D}_s = \{\mathcal{D}_s^k\}_{k=1}^K$ denote K labeled source domains, where each $\mathcal{D}_s^k = \{(\mathbf{x}_i^k, y_i^k)\}$ is drawn from a distribution $\mathbf{P}_k(X, Y)$. The goal is to learn a classifier $f(\cdot)$ that generalizes to an unseen target domain $\mathcal{D}_t \sim P_t(X, Y)$, without access to target

data or domain labels during inference. An overview of the proposed Artifact-aware Prompt Learning for Domain Generalization (APDG) framework is shown in Fig. 1. APDG is built upon a Vision Transformer backbone and introduces prompt-based modeling to facilitate robust generalization. The detailed design of each component is presented in the following subsections.

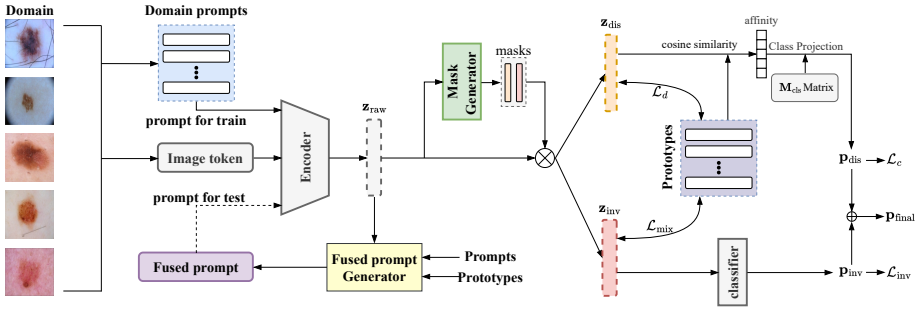

Fig. 1. The framework of APDG.

3.1 Domain-Specific Prompt Modeling

In multi-source dermoscopic image classification, heterogeneous imaging artifacts (e.g., hairs, dark corners, bubbles) introduce domain-specific variations that significantly affect feature distributions. To explicitly model such variations, we treat each artifact type as a distinct source domain and introduce domain-specific visual prompts to encode domain characteristics during training.

Given an input image $\mathbf{x}_i$ from the k-th source domain, the image is first tokenized and embedded into a sequence of visual tokens. For each source domain, a learnable domain-specific prompt $\mathbf{P}_k \in \mathbb{R}^{L\times D}$ is introduced, where $\mathbf{P}_k$ consists of L prompt tokens and D is the embedding dimension. The prompt tokens are concatenated with image tokens and jointly fed into a Vision Transformer (ViT) encoder:

$$\mathbf{z}_{\text{raw}} = f_{\text{ViT}}\big([\mathbf{P}_k; \mathbf{x}_i]\big), \tag{1}$$

where $\mathbf{z}_{\text{raw}} \in \mathbb{R}^{1\times D}$ denotes the encoded feature representation and $[\cdot;\cdot]$ represents token concatenation.

By injecting domain-specific prompts into the feature extraction process, the ViT encoder is encouraged to adapt its attention patterns according to domain-dependent artifact distributions, enabling the model to capture domain characteristics while preserving lesion-relevant semantics.

3.2 Disentanglement and Domain-Specific Prototype Learning

Imaging artifacts in multi-source dermoscopic images are often entangled with lesion semantics, hindering robust generalization. To address this issue, we intro-

duce an explicit feature disentanglement module on top of domain prompt learning, aiming to separate domain-invariant semantic information from domain-specific artifact variations.

Feature Disentanglement via Soft Masking. Given the raw representation $\mathbf{z}_{\text{raw}}$ extracted by the ViT encoder, a lightweight two-layer MLP serves as the Mask Generator to produce a soft mask

$$\mathbf{m} = \sigma(\mathbf{W}_2\, \rho(\mathbf{W}_1 \mathbf{z}_{\text{raw}}))\,, \tag{2}$$

where $\rho(\cdot)$ and $\sigma(\cdot)$ denote ReLU and Sigmoid functions, respectively. Two complementary masks are constructed as $\mathbf{m}_{\text{cls}} = \mathbf{m}$ and $\mathbf{m}_{\text{dis}} = 1 - \mathbf{m}$, yielding disentangled features

$$\mathbf{z}_{\text{inv}} = \mathbf{z}_{\text{raw}} \odot \mathbf{m}_{\text{cls}}, \quad \mathbf{z}_{\text{dis}} = \mathbf{z}_{\text{raw}} \odot \mathbf{m}_{\text{dis}}. \tag{3}$$

The invariant feature $\mathbf{z}_{\text{inv}}$ is used for disease classification with loss

$$\mathcal{L}_{\text{inv}} = \text{CE}\big(g(\mathbf{z}_{\text{inv}}), y_i\big), \tag{4}$$

while $\mathbf{z}_{\text{dis}}$ preserves artifact-related variations.

Domain-Specific Prototype Learning. To explicitly regulate domain-related information, we introduce a prototype matrix $\mathbf{B} \in \mathbb{R}^{CK \times D}$, where each prototype corresponds to a class–domain pair. Two mapping matrices, $\mathbf{M}_{\text{cls}} \in \mathbb{R}^{CK \times C}$ and $\mathbf{M}_{\text{dom}} \in \mathbb{R}^{CK \times K}$, are employed to supervise class and domain information, respectively. The domain classification loss is defined as

$$\mathcal{L}_d = \text{CE}\big(\mathbf{z}_{\text{dis}} \mathbf{B}^\top \mathbf{M}_{\text{dom}}, d_i\big)\,. \tag{5}$$

To retain complementary class-discriminative cues, a prototype-based class loss is introduced:

$$\mathcal{L}_c = \text{CE}\big(\phi(\mathbf{z}_{\text{dis}}, \mathbf{B}^\top)\mathbf{M}_{\text{cls}}, y_i\big)\,, \tag{6}$$

where the affinity function is defined as

$$\phi(\mathbf{z}_{\text{dis}}, \mathbf{B}^\top) = \exp\big(-\beta(1 - \mathbf{z}_{\text{dis}}\mathbf{B}^\top)\big). \tag{7}$$

Domain Confusion and Objective. To suppress residual domain information in the invariant branch, a domain confusion loss is imposed:

$$\mathcal{L}_{\text{mix}} = -\frac{1}{N} \sum_{i=1}^{N} \text{KL}(\text{U} \,\|\, \text{softmax}(\mathbf{logits}_i))\,, \tag{8}$$

where $\text{logits}_i = \phi(\mathbf{z}_{\text{inv}}, \mathbf{B}^\top)\mathbf{M}_{\text{dom}}$. The overall objective is

$$\mathcal{L}_{dp} = \mathcal{L}_{\text{inv}} + \mathcal{L}_{\text{mix}} + \mathcal{L}_d + \lambda \mathcal{L}_c. \tag{9}$$

The domain classification loss $\mathcal{L}_d$ and the domain confusion loss $\mathcal{L}_{\text{mix}}$ form a push–pull mechanism in the feature space: the former encourages $\mathbf{z}_{\text{dis}}$ to explicitly capture inter-domain discrepancies, while the latter suppresses domain-related information in $\mathbf{z}_{\text{inv}}$, thereby stabilizing the disentanglement process. The class loss $\mathcal{L}_c$, introduced with a relatively small weight, prevents the domain-specific features from degenerating into pure domain noise and provides complementary cues for final prediction.

At inference time, predictions from the invariant and domain-specific branches are jointly considered. The invariant branch provides stable semantic discrimination, while the domain-specific branch contributes controlled environment-related cues, leading to improved robustness on unseen target domains.

3.3 Test-Time Prompt Fusion

At inference time, target-domain labels are unavailable, making direct prompt selection infeasible. To address this, we adopt a prototype-guided prompt fusion strategy that adaptively aggregates domain-specific prompts.

Given a test sample $\mathbf{x}_i$, we first obtain its raw feature representation $\mathbf{z}_{\text{raw}} = f(\mathbf{x}_i)$ without introducing any domain prompt. The relevance of each source domain is then estimated by projecting $\mathbf{z}_{\text{raw}}$ onto the learned class–domain prototype space and aggregating responses at the domain level. Specifically, the fusion weights are computed as

$$\boldsymbol{\alpha} = \text{softmax}\big(\mathbf{z}_{\text{raw}}\mathbf{B}^{\top}\mathbf{M}_{\text{dom}} \,/\, \tau\big)\,, \tag{10}$$

where τ is a temperature parameter.

Based on the obtained weights $\boldsymbol{\alpha}$, the fused prompt is constructed as

$$\mathbf{P}_{\text{fused}} = \sum_{k=1}^{K} \alpha_k \mathbf{P}_k. \tag{11}$$

The fused prompt is concatenated with the input image and fed into the ViT encoder for final prediction. This prototype-guided fusion enables adaptive utilization of source-domain knowledge without requiring target-domain information, improving robustness to unseen domain shifts.

4 Experiments

4.1 Task Definition and Datasets

We study binary melanoma classification under artifact-induced domain shifts, with the goal of distinguishing malignant melanoma from benign skin lesions across heterogeneous imaging conditions. In clinical practice, dermoscopic images are often affected by diverse acquisition artifacts, which lead to severe distribution shifts and hinder cross-dataset generalization.

Following prior artifact-aware dermoscopic studies, the training data are constructed from ISIC 2019 [3]. Instead of treating it as a single domain, we partition the training set into five artifact-based source domains using the annotations in [1]: *dark corners*, *hair*, *gel bubbles*, *rulers*, and *vascular structures*, with 4609, 8970, 3550, 1158, and 1862 images, respectively. This artifact-driven partition introduces diverse domain shifts and better reflects real-world acquisition variability.

To evaluate cross-dataset generalization, models are tested on four unseen target datasets collected from independent clinical sources, including Derm7pt-Dermoscopic [10], Derm7pt-Clinical [10], PH2 [14], and PAD-UFES-20 [15]. These datasets differ substantially in acquisition devices and imaging protocols, providing a comprehensive evaluation of out-of-distribution generalization. None of the target datasets are observed during training, and no target-domain data or domain labels are available at inference time.

4.2 Experimental Setup

All experiments are conducted using a ViT-Base/16 backbone pretrained on ImageNet. The model is implemented in PyTorch and optimized with AdamW, with a learning rate of 5×10^{-6}, weight decay 1×10^{-5}, and batch size 40.

For domain modeling, four domain-specific visual prompts are introduced for each source domain. The domain-specific visual prompts are randomly initialized and jointly optimized during training. All images are resized to 224×224, and standard data augmentation is applied. The model is trained for 50 epochs, and the checkpoint with the best validation performance is selected for evaluation. All experiments are conducted on a single NVIDIA RTX 4090D GPU. During inference, test-time prompt fusion is applied to adaptively aggregate domain-specific knowledge.

4.3 Comparison with State-of-the-Art Methods

We compare the proposed method with representative domain generalization approaches, including ERM, DRO, SelfReg, MMLD, and recent prompt-based methods such as DoPrompt, EPVT, and PLDG.

Table 1 reports the ROC-AUC results on four unseen target datasets. ERM exhibits a clear performance degradation under cross-dataset evaluation, highlighting its limited robustness to severe distribution shifts. In contrast, the proposed method achieves consistently strong performance across all unseen datasets, obtaining the best results on Derm7pt-Dermoscopic, Derm7pt-Clinical, and PAD-UFES-20, while remaining highly competitive on PH2. Overall, our method achieves an average ROC-AUC of 85.77%, outperforming the second-best method by 1.34%. These results demonstrate the effectiveness of explicitly modeling domain-specific artifacts for robust cross-dataset dermoscopic image classification.

Table 1. Comparison with existing domain generalization methods on OOD datasets (ROC-AUC, %).

Method	DM7_D	DM7_C	PAD	PH2	Average
ERM	81.42	69.81	77.96	91.50	80.17
DRO [17]	82.55	72.86	80.02	84.97	80.10
MMLD [13]	79.60	69.00	88.80	81.30	79.68
DoPrompt [23]	82.38	71.61	83.81	91.33	82.06
SelfReg [11]	82.49	74.82	79.93	91.03	82.07
EPVT [21]	83.25	74.52	87.41	92.53	84.43
PLDG [20]	83.69	72.03	**89.92**	89.09	83.68
Ours	**83.89**	**78.34**	87.55	**93.28**	**85.77**

Table 2. Ablation study on different components (ROC-AUC, %).

Method	DM7_D	DM7_C	PAD	PH2	Average
Baseline	81.42	69.81	77.96	91.50	80.17
+DP	83.19	76.28	82.29	90.68	83.11
+DP+FD	83.21	75.73	81.02	90.28	82.56
+DP+FD+PL	83.36	76.12	84.76	91.25	83.87
+DP+FD+PL+PF	**83.89**	**78.34**	**87.55**	**93.28**	**85.77**

4.4 Ablation Study

We conduct ablation studies by progressively introducing domain-specific prompts (DP), soft-mask-based feature disentanglement (FD), domain-specific prototype learning (PL), and test-time prompt fusion (PF). The results are summarized in Table 2.

Starting from the ERM baseline, incorporating domain-specific prompts yields a clear improvement, validating the benefit of explicit domain modeling. Adding feature disentanglement alone slightly degrades performance. This occurs because FD and PL are designed as a unified, synergistic mechanism; applying FD alone without the auxiliary supervision signals from PL leads to unstable representations. By further introducing prototype-guided learning, performance is significantly improved, indicating that prototype-level supervision stabilizes the disentanglement process. Combining all components achieves the best overall performance.

5 Conclusion

In this work, we addressed the domain generalization challenge in dermoscopic image classification from an artifact-aware perspective. By explicitly modeling common imaging artifacts as multiple source domains, we proposed a unified prompt-based framework that combines domain-specific prompts, soft-mask-based feature disentanglement, and domain-specific prototype learning to jointly capture domain-related variations while preserving lesion-relevant semantics.

Furthermore, a test-time prompt fusion strategy was introduced to enable adaptive utilization of learned domain knowledge without requiring target do-

main information. Extensive experiments on multiple dermoscopic and clinical datasets demonstrated that the proposed method achieves improved robustness and generalization under severe artifact-induced domain shifts.

These results suggest that explicitly modeling artifact-driven domain characteristics, rather than enforcing strict domain invariance, provides a promising direction for reliable deployment of dermoscopic image analysis models in real-world clinical scenarios, and can readily extend to multi-center settings using hospital indices as domain identifiers.

Acknowledgments. This work is supported by the National Natural Science Foundation of China (Grant No. 62476015).

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

References

1. Bissoto, A., Barata, C., Valle, E., Avila, S.: Artifact-based domain generalization of skin lesion models. In: ECCV Workshops (2022), https://api.semanticscholar.org/CorpusID:251719528
2. Chamarthi, S., Fogelberg, K., Maron, R.C., Brinker, T.J., Niebling, J.: Mitigating the influence of domain shift in skin lesion classification: A benchmark study of unsupervised domain adaptation methods on dermoscopic images. ArXiv **abs/2310.03432** (2023), https://api.semanticscholar.org/CorpusID:263671940
3. Combalia, M., Codella, N.C.F., Rotemberg, V.M., Carrera, C., Dusza, S.W., Gutman, D., Helba, B., Kittler, H., Kurtansky, N.R., Liopyris, K., Marchetti, M.A., Podlipnik, S., Puig, S., Rinner, C., Tschandl, P., Weber, J., Halpern, A., Malvehy, J.: Validation of artificial intelligence prediction models for skin cancer diagnosis using dermoscopy images: the 2019 international skin imaging collaboration grand challenge. The Lancet. Digital health **4**, e330 – e339 (2022), https://api.semanticscholar.org/CorpusID:248338266
4. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V.S.: Domain-adversarial training of neural networks. In: Journal of machine learning research (2015), https://api.semanticscholar.org/CorpusID:2871880
5. Gao, Y., Shi, X., Zhu, Y., Wang, H., Tang, Z., Zhou, X., Li, M., Metaxas, D.N.: Visual prompt tuning for test-time domain adaptation. ArXiv **abs/2210.04831** (2022), https://api.semanticscholar.org/CorpusID:252780157
6. Gulrajani, I., Lopez-Paz, D.: In search of lost domain generalization. ArXiv **abs/2007.01434** (2020), https://api.semanticscholar.org/CorpusID:220347682
7. Gutman, D.A., Codella, N.C.F., Celebi, M.E., Helba, B., Marchetti, M.A., Mishra, N.K., Halpern, A.C.: 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). 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) pp. 168–172 (2016), https://api.semanticscholar.org/CorpusID:10768153
8. Jeong, H.K., Park, C., Henao, R., Kheterpal, M.K.: Deep learning in dermatology: A systematic review of current approaches, outcomes, and limitations. JID Innovations **3** (2022), https://api.semanticscholar.org/CorpusID:251777433

9. Jia, M., Tang, L., Chen, B.C., Cardie, C., Belongie, S.J., Hariharan, B., Lim, S.N.: Visual prompt tuning. ArXiv **abs/2203.12119** (2022), https://api.semanticscholar.org/CorpusID:247618727
10. Kawahara, J., Daneshvar, S., Argenziano, G., Hamarneh, G.: Seven-point checklist and skin lesion classification using multitask multimodal neural nets. IEEE Journal of Biomedical and Health Informatics **23**, 538–546 (2019), https://api.semanticscholar.org/CorpusID:51614248
11. Kim, D., Park, S., Kim, J., Lee, J.: Selfreg: Self-supervised contrastive regularization for domain generalization. 2021 IEEE/CVF International Conference on Computer Vision (ICCV) pp. 9599–9608 (2021), https://api.semanticscholar.org/CorpusID:233307250
12. Lin, M., Weng, N., Mikolaj, K.W., Bashir, Z., Svendsen, M.B.S., Tolsgaard, M.G., Christensen, A.N., Feragen, A.: Shortcut learning in medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (2024), https://api.semanticscholar.org/CorpusID:268356676
13. Matsuura, T., Harada, T.: Domain generalization using a mixture of multiple latent domains. In: AAAI Conference on Artificial Intelligence (2019), https://api.semanticscholar.org/CorpusID:208138410
14. Mendonça, T., Celebi, M., Mendonca, T., Marques, J.: Ph2: A public database for the analysis of dermoscopic images. Dermoscopy image analysis **2** (2015)
15. Pacheco, A.G.C., Lima, G.R., da Silva Salomão, A., Krohling, B., Biral, I.P., de Angelo, G.G., Jr, F.C.A., Esgario, J.G.M., Simora, A.C., Castro, P.B.C., Rodrigues, F.B., Frasson, P.H.L., Krohling, R.A., Knidel, H., Santos, M.C.S., do Espírito Santo, R.B., Macedo, T.L., Canuto, T.R.P., de Barros, L.F.: Pad-ufes-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones. Data in Brief **32** (2020), https://api.semanticscholar.org/CorpusID:220425558
16. Ramé, A., Dancette, C., Cord, M.: Fishr: Invariant gradient variances for out-of-distribution generalization. In: International Conference on Machine Learning (2021), https://api.semanticscholar.org/CorpusID:237431405
17. Sagawa, S., Koh, P.W., Hashimoto, T.B., Liang, P.: Distributionally robust neural networks. In: International Conference on Learning Representations (2020), https://api.semanticscholar.org/CorpusID:213662188
18. Wang, D., Shelhamer, E., Liu, S., Olshausen, B.A., Darrell, T.: Tent: Fully test-time adaptation by entropy minimization. In: International Conference on Learning Representations (2021), https://api.semanticscholar.org/CorpusID:232278031
19. Wang, J., Lan, C., Liu, C., Ouyang, Y., Qin, T.: Generalizing to unseen domains: A survey on domain generalization. IEEE Transactions on Knowledge and Data Engineering **35**, 8052–8072 (2021), https://api.semanticscholar.org/CorpusID:232110832
20. Yan, S., Liu, C., Yu, Z., Ju, L., Mahapatra, D., Betz-Stablein, B., Mar, V., Janda, M., Soyer, P., Ge, Z.: Prompt-driven latent domain generalization for medical image classification. ArXiv **abs/2401.03002** (2024), https://api.semanticscholar.org/CorpusID:266844531
21. Yan, S., Liu, C.C., Yu, Z., Ju, L., Mahapatrainst, D., Mar, V.J., Janda, M., Soyer, P., Ge, Z.: Epvt: Environment-aware prompt vision transformer for domain generalization in skin lesion recognition. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (2023), https://api.semanticscholar.org/CorpusID:257921678

22. Yoon, J.S., Oh, K., Shin, Y., Mazurowski, M.A., Suk, H.I.: Domain generalization for medical image analysis: A review. Proceedings of the IEEE **112**, 1583–1609 (2023), https://api.semanticscholar.org/CorpusID:274555557
23. Zheng, Z., Yue, X., Wang, K., You, Y.: Prompt vision transformer for domain generalization. ArXiv **abs/2208.08914** (2022), https://api.semanticscholar.org/CorpusID:251643490
24. Zhou, K., Liu, Z., Qiao, Y., Xiang, T., Loy, C.C.: Domain generalization: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**, 4396–4415 (2021), https://api.semanticscholar.org/CorpusID:232104764
25. Zhou, K., Yang, J., Loy, C.C., Liu, Z.: Learning to prompt for vision-language models. International Journal of Computer Vision **130**, 2337 – 2348 (2021), https://api.semanticscholar.org/CorpusID:237386023