


MedPro-DG: Domain-Aware Masked Contrastive Prompt Learning of Institution Generalization for Outcome Prediction

Rongfang Wang¹ , JiaSheng Chen¹, Xinlong Zhang², Jing Wang², Hui Liu³, Zhiguo Zhou³, and Kai Wang⁴

¹ School of Artificial Intelligence, Xidian University, China
rfwang@xidian.edu.cn

² Department of Radiation Oncology, UTSW, United States of America

³ Department of Biostatistics Data Science, KUMC, United States of America

⁴ Department of Radiation Oncology, UMMC, United States of America

Abstract. Accurate outcome prediction for head and neck cancer is critical but remains challenging due to domain shifts across multi-institutional imaging datasets. Existing domain generalization (DG) methods focus on visual features while overlooking clinical domain-invariant information. To address this gap, we propose MedPro-DG, a novel prompt learning framework that integrates CT imaging with clinical variables using domain-aware masked contrastive prompt learning. Our method can effectively mitigate domain shifts by aligning cross-modal features with domain-invariant clinical semantics. Extensive experiments conducted across six medical centers demonstrate the superiority of MedPro-DG, which outperforms state-of-the-art DG methods by 1.35% in AUC and 4.06% in ACC on average. Ablation studies further reveal that our prompt learning can capture clinically domain-invariant features, highlighting their diagnostic relevance. This work pioneers domain-invariant vision-language fusion for medical domain generalization, providing an available and effective solution for multi-center collaborative diagnosis.

Keywords: Domain Generalization · Clinical Information · Prompt Learning · Domain Contrastive Learning · Head and Neck Cancer.

1 Introduction

Head and neck (H&N) cancer is a group of tumors originating from squamous cells on the surface of mucous membranes in the oral cavity, sinuses, pharynx or larynx [2]. H&N cancers account for approximately 5% of all malignancies, with more than 500,000 new cases diagnosed worldwide each year [22]. Primary treatment modalities include surgery, radiation therapy, and chemotherapy, while advanced stage disease often requires a multimodal therapeutic approach [5]. With the advancement of deep learning, researchers have proposed various prediagnostic models to assess the degree of pathological differentiation in H&N cancers [7,10,24], aiming to assist clinicians in diagnosis. However, existing

methods typically rely on data from a single medical institution (or domain) for model training and evaluation. Although these models achieve satisfactory performance on test sets from the same institution, their generalizability decreases significantly when applied to data from different institutions [11]. This limitation arises because medical data are collected from various hospitals and medical devices, resulting in substantial domain variations (e.g., differences in imaging equipment, acquisition protocols, or patient demographics). Such variations can lead to poor performance of deep learning-based models when deployed in new clinical settings.

Domain Generalization (DG) methods [19,21,23,26] have been developed to address these challenges by training models to extract domain-invariant features, thus minimizing discrepancies across domains. Traditional DG approaches often focus on image data, employing techniques such as data augmentation[25], domain-adversarial training [4], and meta-learning [13]. Data augmentation enhances model robustness by artificially expanding the training dataset with transformed samples. Domain-adversarial training encourages the model to learn features that are indistinguishable across domains. Meta-learning simulates domain shifts during training to improve generalization to unseen domains. However, these methods frequently overlook the multimodal nature of medical data, particularly the integration of clinical information, which is generally more consistent across institutions due to standardized medical terminologies and guidelines.

In this paper, we propose MedPro-DG, a vision-language framework for domain generalization in outcome prediction of H&N cancer. The framework synergizes CT imaging with clinical text through domain-aware masked contrastive prompt learning, effectively integrating image data with clinical text information. Our framework tackles cross-domain medical prediction through two key innovations: an Attention-Augmented Visual Prompt (AAVP) that integrates spatial attention from CT imaging with learnable text prompts to fuse domain-invariant clinical semantics, and a Domain-Masked Contrastive Loss (DMCL) enforcing cross-domain alignment of same-class clinical text embeddings while repelling different-class pairs. Specifically, AAVP encodes tumor location attention maps from ResNet50 [8] layer4 to dynamically modulate clinical text prompts, bridging imaging and clinical narratives in a domain-robust manner. DMCL further suppresses domain-specific biases by redefining positive pairs as same-diagnosis samples across institutions and negative pairs as different-diagnosis samples, ensuring text embeddings capture pathology-centric semantics invariant to shifts in imaging protocols or device vendors. Extensive experiments across six centers demonstrate the superiority of MedPro-DG, achieving the highest ACC of 83.70% and AUC of 74.34%. Ablation studies reveal individual contributions.

2 Method

2.1 Overview

Let $\mathcal{D} = \{D_1, \dots, D_6\}$ denote multi-domain data from six medical centers, where each domain D_i contains CT images, clinical variables, and binary labels (lo-coregional recurrence or normal). Each sample is represented as $(\mathbf{x}_{CT}, \mathbf{x}_{Clin}, y)$, with \mathbf{x}_{CT} and \mathbf{x}_{Clin} denoting the imaging and clinical features respectively. We aim to learn a domain generalizable model from \mathcal{D} without accessing the data from the target domain during training. Our framework aims to enhance domain generalization for medical image analysis using clinical text information and a novel attention-augmented prompt learning strategy. The core idea is to freeze all backbone networks (image and text encoders) while only training learnable prompts and a domain-aware contrastive loss module. This design reduces overfitting risks and ensures efficient adaptation to multi-domain clinical scenarios. The overall framework of MedPro-DG is shown in Fig.1.

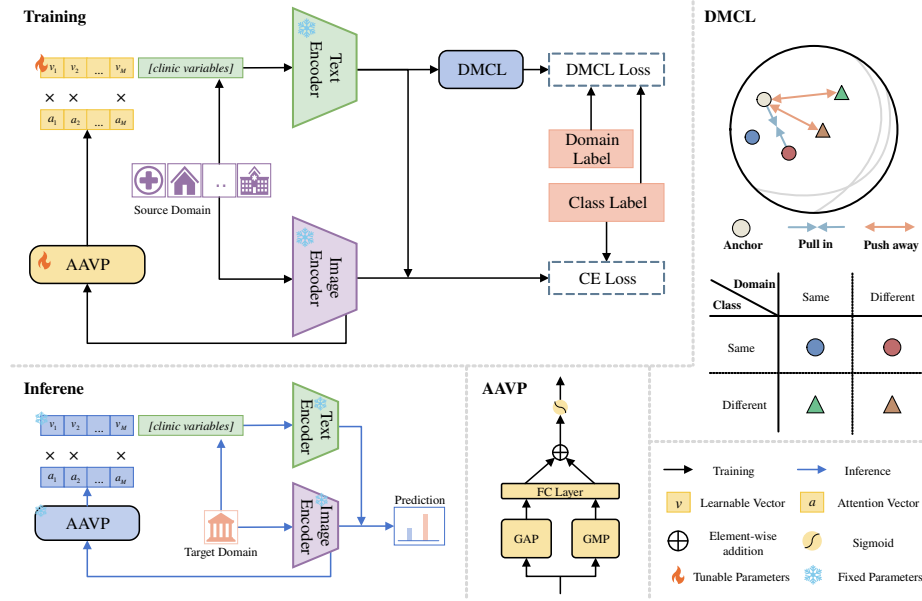


Fig. 1. The overall framework of MedPro-DG. It learns generalizable representations through Attention-Augmented Visual Prompt (AAVP) and Domain-Masked Contrastive Learning (DMCL). GAP: global average pooling. GMP: global max pooling.

2.2 Attention-Augmented Visual Prompt

Medical imaging domains (for example, different CT scanners) often exhibit local texture variations [26]. Our Attention-Augmented Visual Prompt (AAVP)

explicitly models cross-domain spatial dependencies through dual-path attention. AAVP’s attention mechanism automatically focuses on pathology-relevant regions while ignoring scanner-specific artifacts. We use ResNet50 [8] as the image encoder and CLIP[16] text encoder to extract specific clinical information. Note that the training parameters are frozen for the encoders of the above two branches. Concretely, instead of using ‘a photo of a [class]’ as a context, we encode clinical variables by concatenating them into a simple sentence, e.g., "57, Oropharynx, T2, N0, radiation, positive". Furthermore, following [27,28], we introduce $M = 16$ learnable context vectors $P_{\text{learnable}} = \{v_1, v_2, \dots, v_M\}$, each of which has the same dimension as the clinic embeddings. In the last layer of ResNet50 [8], layer4, we extracted its features as input to our AAVP:

$$F_{\text{Layer4}} = \text{ResNet50}_{\text{frozen}}(\mathbf{x}_{\text{CT}}) \quad (1)$$

Then, we derive spatial attention through a dual-path adapter [14]:

$$\begin{cases} F_{\text{max}} = W_1 \cdot \text{MaxPool}(F_{\text{Layer4}}) + b_1 \\ F_{\text{avg}} = W_2 \cdot \text{AvgPool}(F_{\text{Layer4}}) + b_2 \end{cases} \quad (2)$$

$$\alpha = \sigma(F_{\text{max}} + F_{\text{avg}}) \quad (3)$$

$$T = \text{CLIP}_{\text{text}}[\text{Concate}[\alpha \odot P_{\text{learnable}}, P_{\text{clinic}}]] \quad (4)$$

where W_1, W_2 are trainable weights, σ denotes sigmoid activation, P_{clinic} is from clinic information, see in Section 3.1. Each learnable prompt $P_{\text{learnable}}$ is initialized with a Gaussian distribution parameterized by $\mathcal{N}(\mu_c, \sigma_c^2)$. \odot means element-wise multiplication.

2.3 Domain-Masked Contrastive Learning

Clinical narratives inherently encode disease-specific semantics that are less sensitive to domain shifts, whereas imaging features often entangle pathological patterns with scan artifacts. Using this property, domain-masked contrast learning (DMCL) explicitly aligns text embeddings through domain-aware constraints to disentangle domain-invariant clinical concepts. Let there be N samples in the batch:

$$N = \sum N_{D_i} \quad (5)$$

where N_{D_i} denotes the number of samples in the source domain i , and the clinical text feature vector for each sample is obtained through the AAVP.

Following the standard practice in supervised contrastive learning, we compute the cosine similarity matrix S between all text embeddings. The key innovation lies in our domain-aware masking strategy for selecting positive and negative pairs. Eq.7 is the loss function based on supervised contrastive learning:

$$S_{ij} = T_i \cdot T_j^\top \quad (6)$$

$$\mathcal{L}_{\text{sup}} = - \sum_{i \in I} \frac{1}{|\mathcal{P}_i|} \sum_{p \in \mathcal{P}_i} \log \frac{\exp(S_{ip}/\tau)}{\sum_{a \in \mathcal{A}_i} \exp(S_{ia}/\tau)} \quad (7)$$

where positive samples \mathcal{P}_i denote all samples in the same class as anchor i (whether from the same domain or not), and negative samples \mathcal{A}_i denote all samples except the anchor (including samples in the same class with different domains).

As shown in Eq.7, all intra-class samples are treated as positives. However, our approach introduces a domain-masked contrastive loss, which incorporates an additional domain label d_i . Domain labels are temporarily generated before training. We incorporate both domain and category labels to define positive and negative sample pairs. Specifically, for the i -th sample:

$$\mathcal{N}_i = \{j \mid y_j \neq y_i\} \quad (8)$$

$$\mathcal{P}_i = \{j \mid y_j = y_i \text{ and } d_j \neq d_i\} \quad (9)$$

This allows us to selectively pair samples from different domains, thus enforcing domain-invariant feature learning. The details are shown in Eq.10.

$$\mathcal{L}_{\text{DMCL}} = -\frac{1}{N} \sum_{i=1}^N \frac{1}{|\mathcal{P}_i|} \sum_{j \in \mathcal{P}_i} \log \frac{\exp(S_{ij}/\tau)}{\underbrace{\sum_{p \in \mathcal{P}_i} \exp(S_{ip}/\tau)}_{\text{same class, different domains}} + \underbrace{\sum_{k \in \mathcal{N}_i} \exp(S_{ik}/\tau)}_{\text{different class}}} \quad (10)$$

This loss is combined with classification loss, where we set $\lambda = 1.0$.

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CE}} + \lambda \mathcal{L}_{\text{DMCL}}, \quad (11)$$

3 Experiments and Results

3.1 Dataset

Our experiments utilized datasets from six medical centers, each of which provided unique clinical and imaging data. These datasets are obtained from three sources: The Cancer Imaging Archive (TCIA) [3], the "H&N1" dataset from Maastricht University Medical Center (MUMC) [1], and a dataset collected by the University of Texas Southwestern Medical Center (UTSW) [24]. The TCIA datasets [3] include clinical and PET/CT imaging data from 298 patients treated at four medical centers between April 2006 and November 2014: CHUM, CHUS, HGJ, HMR. The MUMC dataset [1], sourced from the "H&N1" dataset associated with a study published in Nature Communications, comprises clinical data and computed tomography (CT) scans of 137 patients who underwent radiation therapy at the Maastricht University Medical Center in the Netherlands. The UTSW dataset [24] consists of clinical and PET/CT imaging data from 615 patients who received radiotherapy at the University of Texas Southwestern Medical Center in Dallas, Texas, between September 2005 and November 2015.

Since PET imaging is not available at MUMC, we utilized only CT images and clinical data for our study. To ensure a sufficient duration of follow-up,

Table 1. The number of samples from six different medical centers. We excluded all samples with a follow-up period of less than one year.

Class	CHUM	CHUS	HGJ	HMR	MUMC	UTSW	\sum_C
LRR	6	14	12	6	13	49	100
Normal	56	79	65	13	92	157	462
\sum_D	62	93	77	19	105	206	562

we excluded all samples with a follow-up period of less than one year. Tab.1 shows the number of samples from each medical center after screening. Clinical variables include: 1) patient’s age at diagnosis (Age), 2) Primary tumor site (Primary Site), 3) Tumor Stage, indicating tumor size and extent (T-stage), 4) Nodal Stage, indicating lymph node involvement (N-stage), 5) type of treatment received (Therapy), and 6) Human Papillomavirus Status, indicating whether the patient tested positive or negative for HPV (HPV Status).

3.2 Experimental Setup

We utilize ResNet-50 as the backbone network for our classification architecture, integrating it with the CLIP [16] text encoder to effectively combine clinical text and imaging data. For all methods, ResNet-50 uses ImageNet-pretrained weights, including our image encoder. The CLIP text encoder uses the officially provided "RN50" pretrained weights. The model is trained over 1600 iterations with a batch size of 80, distributed as 16 per source domain. During training, we save a model checkpoint every 40 iterations. The model checkpoint with the best performance on the validation set is selected for the final evaluation. Stochastic Gradient Descent (SGD) is used as the optimizer, with an initial learning rate of 0.01, a momentum of 0.9, and a weight decay of $5e-4$. The weighting parameter λ in Eq.11 is set to 1.0. All experiments were performed with Python 3.7, PyTorch 1.12.0, and an Nvidia RTX3090 GPU with 24GB of memory.

For evaluation, we adopt the leave-one-domain-out setting [6], a widely used scheme in domain generalization. In this setting, one domain is designated as the target, while the remaining domains are used as source data for model training. The target domain remains unseen during training, and the model is evaluated on this unseen target domain. We report the accuracy (ACC) and area under the curve (AUC). All models in this study are based on ResNet-50 and share identical network architectures. The only exception is DANN, which includes an additional domain classifier. Consequently, detailed calculations of the model parameters and computational complexity are not central to our discussion.

3.3 Comprehensive Comparison with Other Methods

We first compare the performance of our method with the state-of-the-art methods in DG. We choose Baseline(EMR [20]), DANN [4], CORAL [18], VREx [12],

Mixup [25], MLDG [13], GroupDRO [17], RSC [9] and ANDMask [15]. In addition, we have chosen ERM [20] in DG to incorporate our method, as it shows strong competitiveness with many existing DG methods.

As summarized in Fig.2, MedPro-DG achieves an average ACC of 83.70% and AUC of 74.34% in six medical centers under a leave-one-domain-out protocol, demonstrating superior cross-domain generalization capability. Compared to existing domain generalization methods, MedPro-DG outperforms VREx by 4.06% (83.70% vs. 79.64%) and RSC by 5.32% (83.70% vs. 78.38%) in ACC, validating its robustness to clinical text-variability across hospitals, while also surpassing Mixup by 1.35% (74.34% vs. 72.99%) and RSC by 1.96% (74.34% vs. 72.38%) in AUC, indicating more substantial discriminative power for imbalanced outcome prediction tasks (e.g., locoregional recurrence vs. normal).

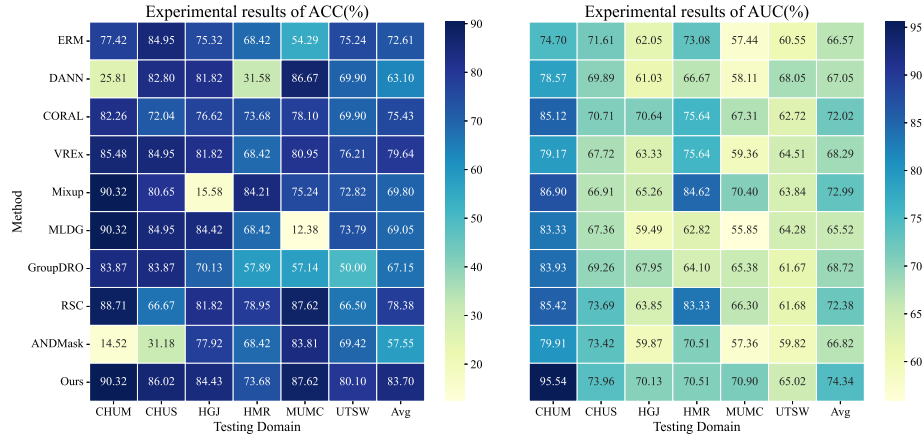


Fig. 2. Heatmap of Domain Generalization Performance: Darker Shades Indicate Higher ACC/AUC, Highlighting Domain-Specific Generalization Challenges.

3.4 Ablation Studies

We conduct the ablation experiments to explore the improvement in model performance by AAVP and DMCL.

AS1: Effect of both AAVP and DMCL: In the model without AAVP and DMCL, we remove them and only reserve the learnable prompt.

AS2: Effect of AAVP: In the model without DMCL, we remove the DMCL loss and use only the CE loss as the total loss.

AS3: Effect of DMCL: In the model without AAVP, we remove the AAVP and reserve the learnable prompt.

AS4: Effect of Domain-Masked in DMCL: In the model without Domain-Masked constraints in DMCL, we use only class labels to construct positive and negative samples, as shown in Eq.7.

Table 2. The improvement of model performance by AAVP and DMCL. **AS4** removing domain-masked constraints and using supervised contrastive learning.

Ablation Studies	AAVP	DMCL			<i>Avg.</i>	
		Domain	Label	Class Label	ACC(%)	AUC(%)
AS1					77.96	69.07
AS2	✓				80.01	68.81
AS3		✓		✓	82.70	70.77
AS4	✓			✓	81.96	70.67
MedPro-DG	✓	✓		✓	83.70	74.34

Table 3. Impact of Domain-Masked Contrastive Loss Weight (λ).

<i>Avg</i>	λ					
	0.1	0.50	0.75	1.0	1.25	1.5
ACC	83.40	82.63	80.61	83.70	81.44	81.45
AUC	71.39	71.07	73.67	74.34	73.01	71.79

The above four factors are analyzed and the experimental results are shown in Tab.2. The baseline configuration (AS1), without AAVP or DMCL, yields 77.96% ACC and 69.07% AUC, reflecting the limitations of naive imaging-text fusion. Enabling AAVP (AS2) increases ACC by 2.05% by paying attention to tumor location, although its isolated use slightly reduces AUC (-0.26%), indicating that spatial guidance alone is insufficient for domain alignment. Activating DMCL (AS3) significantly improves both metrics, achieving 82.70% ACC (+4.74%) and 70.77% AUC (+1.70%), as domain-masked contrastive learning aligns cross-institutional semantics. The complete model synergizes both components, reaching 83.70% ACC and 74.34% AUC, increasing an absolute gain of 5.74% and 5.27% over the baseline (AS1). This synergy highlights the role of AAVP in refining anatomy-aware characteristics and the capacity of DMCL to suppress domain shift, ultimately enhancing prognostic reliability. AS4 removing domain-masked constraints and using standard contrastive learning through Eq.7, underperforming the full DMCL.

Tab.3 analyzes the impact of contrastive weight loss λ in DMCL. Optimal performance is achieved at $\lambda = 1.0$ with 83.70% ACC and 74.34% AUC, balancing classification and alignment between domains. Overweighting contrastive learning degrades ACC, whereas lower values compromise AUC due to insufficient domain invariance. This underscores the necessity of calibrated multimodal alignment for robust generalization.

4 Conclusion

This work alleviates domain shifts in multi-center medical imaging through attention-augmented visual prompts (AAVP) and domain-masked contrastive learning (DMCL). Our framework aligns imaging features with diagnostic semantics by anchoring clinical text as a domain-invariant semantic guide. The

proposed AAVP refines anatomy-aware localization by attention, while DMCL enforces cross-domain consistency by contrasting same-class samples from distinct institutions. We enhance the generalization to unseen data from different domains. Clinically, our framework offers an available and effective solution for multi-center collaboration, enabling reliable prognosis without requiring domain-specific fine-tuning.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (No.62176196), the Key Industry Innovation Chain Project of Shaanxi (No.2024NCZDCYL-05-04).

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

References

1. Aerts, H.J., Velazquez, E.R., Leijenaar, R.T., Parmar, C., Grossmann, P., Carvalho, S., Bussink, J., Monshouwer, R., Haibe-Kains, B., Rietveld, D., et al.: Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature communications* **5**(1), 4006 (2014)
2. Atun, R., Jaffray, D.A., Barton, M.B., Bray, F., Baumann, M., Vikram, B., Hanna, T.P., Knaul, F.M., Lievens, Y., Lui, T.Y., et al.: Expanding global access to radiotherapy. *The lancet oncology* **16**(10), 1153–1186 (2015)
3. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., et al.: The cancer imaging archive (tcia): maintaining and operating a public information repository. *Journal of digital imaging* **26**, 1045–1057 (2013)
4. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M., Lempitsky, V.: Domain-adversarial training of neural networks. *Journal of machine learning research* **17**(59), 1–35 (2016)
5. Grégoire, V., Lefebvre, J.L., Licitra, L., Felip, E.: Squamous cell carcinoma of the head and neck: Ehn-s-esmo-estro clinical practice guidelines for diagnosis, treatment and follow-up. *Annals of oncology* **21**, v184–v186 (2010)
6. Gulrajani, I., Lopez-Paz, D.: In search of lost domain generalization. *arXiv preprint arXiv:2007.01434* (2020)
7. Gupta, P., Malhi, A.K.: Using deep learning to enhance head and neck cancer diagnosis and classification. In: 2018 IEEE international conference on system, computation, automation and networking (icscan). pp. 1–6. IEEE (2018)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
9. Huang, Z., Wang, H., Xing, E.P., Huang, D.: Self-challenging improves cross-domain generalization. In: Computer vision–ECCV 2020: 16th European conference, Glasgow, UK, August 23–28, 2020, proceedings, part II 16. pp. 124–140. Springer (2020)
10. Huynh, B.N., Groendahl, A.R., Tomic, O., Liland, K.H., Knudtsen, I.S., Hoebbers, F., van Elmpt, W., Malinen, E., Dale, E., Futsaether, C.M.: Head and neck cancer treatment outcome prediction: A comparison between machine learning with conventional radiomics features and deep learning radiomics. *Frontiers in medicine* **10**, 1217037 (2023)

11. Korevaar, S., Tennakoon, R., Bab-Hadiashar, A.: Failure to achieve domain invariance with domain generalization algorithms: An analysis in medical imaging. *IEEE Access* **11**, 39351–39372 (2023)
12. Krueger, D., Caballero, E., Jacobsen, J.H., Zhang, A., Binas, J., Zhang, D., Le Priol, R., Courville, A.: Out-of-distribution generalization via risk extrapolation (rex). In: *International conference on machine learning*. pp. 5815–5826. PMLR (2021)
13. Li, D., Yang, Y., Song, Y.Z., Hospedales, T.: Learning to generalize: Meta-learning for domain generalization. In: *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI 2018)*. pp. 3490–3497. AAAI press (2018)
14. Nirthika, R., Manivannan, S., Ramanan, A., Wang, R.: Pooling in convolutional neural networks for medical image analysis: a survey and an empirical study. *Neural Computing and Applications* **34**(7), 5321–5347 (2022)
15. Parascandolo, G., Neitz, A., Orvieto, A., Gresele, L., Schölkopf, B.: Learning explanations that are hard to vary. *arXiv preprint arXiv:2009.00329* (2020)
16. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: *International conference on machine learning*. pp. 8748–8763. PMLR (2021)
17. Sagawa, S., Koh, P.W., Hashimoto, T.B., Liang, P.: Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization. *arXiv preprint arXiv:1911.08731* (2019)
18. Sun, B., Saenko, K.: Deep coral: Correlation alignment for deep domain adaptation. In: *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8–10 and 15–16, 2016, Proceedings, Part III 14*. pp. 443–450. Springer (2016)
19. Teterwak, P., Saito, K., Tsiligkaridis, T., Saenko, K., Plummer, B.A.: Erm++: An improved baseline for domain generalization. In: *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. pp. 8525–8535. IEEE (2025)
20. Vapnik, V.: *The nature of statistical learning theory*. Springer science & business media (2013)
21. Vedantam, R., Lopez-Paz, D., Schwab, D.J.: An empirical investigation of domain generalization with empirical risk minimizers. *Advances in neural information processing systems* **34**, 28131–28143 (2021)
22. Vermorken, J.B.: Medical treatment in head and neck cancer. *Annals of oncology* **16**, ii258–ii264 (2005)
23. Wang, J., Lan, C., Liu, C., Ouyang, Y., Qin, T., Lu, W., Chen, Y., Zeng, W., Philip, S.Y.: Generalizing to unseen domains: A survey on domain generalization. *IEEE transactions on knowledge and data engineering* **35**(8), 8052–8072 (2022)
24. Wang, R., Guo, J., Zhou, Z., Wang, K., Gou, S., Xu, R., Sher, D., Wang, J.: Locoregional recurrence prediction in head and neck cancer based on multi-modality and multi-view feature expansion. *Physics in Medicine & Biology* **67**(12), 125004 (2022)
25. Zhang, H.: mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412* (2017)
26. Zhou, K., Liu, Z., Qiao, Y., Xiang, T., Loy, C.C.: Domain generalization: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**(4), 4396–4415 (2022)
27. Zhou, K., Yang, J., Loy, C.C., Liu, Z.: Conditional prompt learning for vision-language models. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 16816–16825 (2022)

28. Zhou, K., Yang, J., Loy, C.C., Liu, Z.: Learning to prompt for vision-language models. *International Journal of Computer Vision* **130**(9), 2337–2348 (2022)