This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

DualCaMIL: Dual-Level Causal Multi-Instance Learning for Patient-Level Diagnosis in Reflectance Confocal Microscopy

Changxin Wang[1†], Lifang Chen[1,2†*], and Yunmin Zou[3,4]

[1] School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, Jiangsu, China
[2] Engineering Research Center of Intelligent Technology for Healthcare, Ministry of Education, China
[3] Department of Dermatology, Jiangnan University, Wuxi, Jiangsu, China
[4] School of Medicine, Southeast University, Nanjing, Jiangsu, China
6243112050@stu.jiangnan.edu.cn, chenlifang@jiangnan.edu.cn

Abstract. Reflectance confocal microscopy (RCM) enables non-invasive in vivo imaging for basal cell carcinoma (BCC) diagnosis. Clinically, multiple RCM images are typically acquired per patient, while diagnostic labels are only available at the patient level, framing this task as a patient-level multiple instance learning (MIL) problem. Nevertheless, instance-level imaging artifacts, patient-specific macro-patterns, and modality-specific styles often act as confounders, introducing spurious correlations. We propose DualCaMIL, a dual-level causality-driven MIL framework that disentangles non-causal factors at both micro and macro scales. At the instance level, we introduce Cross-modal Microscopic Causal Disentanglement (CMCD), which leverages histopathology as a structural reference to isolate pathology-relevant content from imaging interference. At the patient level, we propose the Patient-aware Macro-level Causal Module (PMCM), which weakens shared non-causal structure among patient instances and prevents dependence on patient-specific contextual biases. Experiments on a public multicenter patient-level RCM dataset show that DualCaMIL outperforms representative protocol-compatible MIL baselines, with an AUC of 0.9447. Counterfactual stress testing further suggests robustness to non-causal factors, including imaging style variations and patient-specific constant patterns, while maintaining sensitivity to lesion-relevant signals. Code is available at `https://github.com/Taffee9/DualCaMIL`.

Keywords: Reflectance Confocal Microscopy · Multi-instance Learning · Causal Learning · Cross-modal Learning.

1 Introduction

Basal cell carcinoma (BCC) is among the most prevalent malignant skin cancers. Its heterogeneous clinical manifestations make accurate early diagnosis essential

† Equal contribution. * Corresponding author.

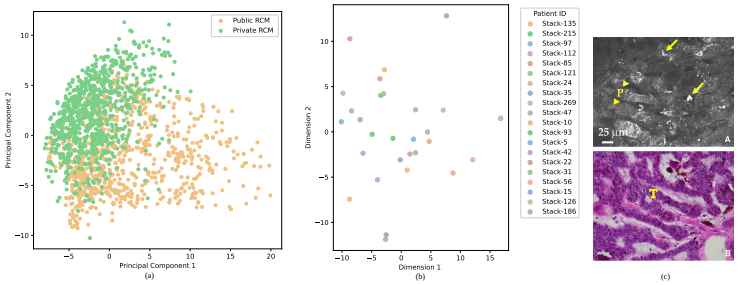

Fig. 1. Patient-level MIL in RCM: confounders and structural alignment. (a) Instance-level feature shift across datasets. PCA visualization reveals distributional differences between public and private RCM data, reflecting non-pathological style variation. (b) Qualitative patient-specific clustering under naive MIL aggregation. Each point denotes one patient-level embedding; the dispersion suggests that average pooling may amplify intra-patient macro-structural patterns over lesion-relevant evidence. (c) Representative RCM–histopathology correspondence in BCC [5]. The RCM image shows pigmented cord-like tumor strands with peripheral palisading (P) and low-reflectance clefts (arrow), with bright oval/stellate cells; the matched H&E section exhibits concordant cord-like growth (T), confirming structural agreement.

to reduce uncertainty, avoid unnecessary biopsies, and optimize treatment decisions [1]. In clinical practice, histopathological examination provides definitive morphological evidence at both cellular and tissue levels and has long been regarded as the gold standard for skin tumor diagnosis [2]. Despite its reliability, histopathology is invasive, limits real-time and longitudinal assessment, and imposes a significant patient burden. RCM enables non-invasive, near-cellular in vivo imaging, revealing horizontal skin architecture and serving as a promising "virtual biopsy" technique [3, 4]. In BCC, RCM demonstrates strong structural concordance with histopathology, enabling the in vivo characterization of key tumor-related architectural features [4, 5]. Clinically, multiple RCM images are acquired per patient, while labels are assigned at the patient level, aligning with the MIL paradigm [6].

From a causality-inspired perspective, patient-level MIL for RCM is inherently affected by multi-level confounders. Here, patient instances refer to the multiple RCM images or patches that form one patient bag. At the instance level, features encode pathological signals together with noise and device-specific imaging styles, inducing dataset shifts (Fig. 1(a)) and spurious correlations with labels. At the patient level, naive aggregation can amplify stable yet disease-irrelevant macro-structural patterns, such as site, texture, pigmentation, or tissue context (Fig. 1(b)), leading to patient-specific clustering and compromised cross-patient stability. These observations suggest that enforcing invariance at a single level is insufficient to ensure robust diagnosis.

Prior studies address confounding in medical imaging through explicit causal modeling and invariance-based learning. Structural Causal Models (SCMs) formalize the image generation process and disentangle anatomical structure from

pathology via intervention or counterfactual reasoning: Castro et al. apply causal interventions to separate anatomy and pathology [7], and Pawlowski et al. develop deep structural causal models for counterfactual lesion synthesis [8]. Complementary approaches mitigate scanner-, site-, and protocol-induced distribution shifts using Invariant Risk Minimization (IRM) or intervention-based objectives [9–11]. Within MIL, Wu et al. incorporate causal modeling into attention aggregation to reduce selection bias and spurious correlations[12]. Nevertheless, existing patient-level MIL frameworks seldom model instance- and patient-level confounders within a unified formulation. Moreover, although RCM and histopathology exhibit stable structural correspondence [5], this cross-modal causal reference has not been systematically integrated into patient-level diagnostic modeling. Ignoring intra-patient structural regularities may further cause invariance enforcement to suppress discriminative pathology, thereby disrupting the balance between robustness and lesion-specific sensitivity.

We propose DualCaMIL, a dual-level causality-driven MIL framework for patient-level BCC diagnosis from RCM images. It jointly models and suppresses non-causal factors at the instance and patient levels. At the instance level, the Cross-modal guided micro-level causal disentanglement (CMCD) module mitigates modality-specific interference. At the patient level, the Patient-aware macro-level causal modeling (PMCM) module attenuates consistent yet disease-irrelevant intra-patient structural biases during aggregation. By jointly constraining both confounders without additional annotations, DualCaMIL enhances diagnostic stability under the evaluated patient-level RCM protocol.

The main contributions of this work are threefold:

1. We introduce the CMCD module for micro-level causal disentanglement. It uses histopathology-guided constraints to separate disease-related structures from modality interference and improve robustness to style variations and domain shifts.
2. We propose the PMCM module to capture and suppress patient-shared structural components unrelated to disease, reducing their interference during MIL aggregation and improving cross-patient stability.
3. We propose DualCaMIL, a dual-level causality-driven MIL framework for BCC diagnosis from RCM images, along with a patient-level counterfactual stress-testing framework to evaluate robustness against non-causal perturbations and reliance on lesion-specific evidence.

2 Method

2.1 Overview

We propose DualCaMIL, a dual-level causality-driven framework (Fig. 2(a)). The framework comprises two core components: 1) The CMCD module (Fig. 2(b)) uses histopathology guidance to disentangle pathology-relevant signals from modality-specific confounders at the instance level; 2) The PMCM module (Fig. 2(c)) attenuates non-causal intra-patient structural patterns at the macro level to improve cross-patient stability.

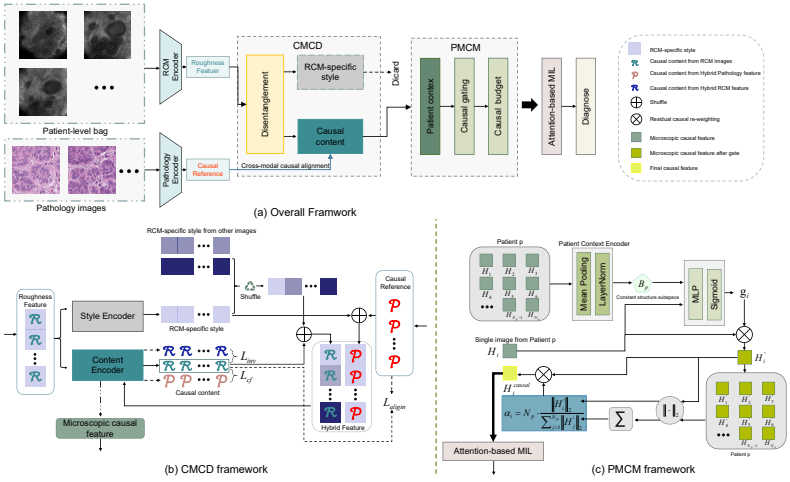

Fig. 2. (a) Overview of the model framework: Multiple RCM images from a patient are disentangled via CMCD, refined through PMCM, and aggregated by Attention-MIL for diagnosis. (b) CMCD framework: Histopathology-guided disentanglement of RCM features into content (c_{rcm}) and style (s_{rcm}), using counterfactual constraints to encourage modality-invariant pathological representations. (c) PMCM schematizes covariance-based patient-context estimation by eigendecomposition, followed by direction-aware modulation and causal budget competition; it does not introduce an additional context encoder.

2.2 Cross-modal guided micro-level causal disentanglement

RCM images encode cellular pathological structures but also contain speckle, attenuation, and device-related style variations. These factors are entangled with pathological signals, making separation difficult under standard training. We therefore propose the CMCD module. In this module, a ConvNeXt V2 backbone extracts feature $\mathbf{z}$, which is then decomposed into pathology-relevant content c_{rcm} and RCM-specific style s_{rcm} representing modality variations.

We impose orthogonality constraints between content and style to reduce information leakage, together with a feature-level reconstruction loss to preserve reversibility and information integrity. Since RCM-only disentanglement may retain modality-specific bias, we introduce semantically paired histopathological features c_{path} to guide c_{rcm} via cosine similarity, providing morphological-level alignment:

$$\mathcal{L}_{\text{align}} = \|c_{rcm} - c_{path}\|_2^2 \tag{1}$$

The pairing is class- and semantics-matched rather than patient- or region-registered: histopathology tiles act as structural teachers and do not require spatial correspondence with RCM images. This alignment is grounded in biological and structural evidence. Agero et al. demonstrated that RCM functions as in vivo histology, where key BCC structures—including tumor nests and palisading—exhibit high concordance with histopathology[5]. As illustrated in Fig.

1(c), hyporefractile tumor nests and clefts in RCM correspond to analogous cord-like or nodular patterns in H&E sections, preserving related spatial morphology and tumor–stroma boundaries across subtypes. Despite differing sectioning orientations, the nearly isotropic 3D architecture of tumor nests maintains comparable topology across modalities, and horizontal sections from Mohs surgery further confirm agreement in boundary delineation[13]. At the representational level, virtual staining and generative modeling studies by Milind Rajadhyaksha and David Gareau[14, 15] reveal a substantially shared and transformable latent embedding space between RCM and H&E representations. Low-quality or semantically inconsistent pathology tiles may weaken the teacher signal; semantic matching, reconstruction, orthogonality, and intervention consistency mitigate but do not eliminate this risk.

Building on this foundation, we introduce a feature-level weak causal intervention to enhance content robustness. Specifically, during training, we randomly shuffle RCM style features within each batch and recombine them with original content features to construct hybrid representations $\hat{z} = g(c_{rcm}, s^{p}_{rcm})$, where p represents the random permutation function. The re-extracted content c_{rcm} is constrained to match the original content:

$$\mathcal{L}_{\text{inv}} = \|\tilde{c}_{rcm} - c_{rcm}\|_2^2 \tag{2}$$

For cross-modal pairs, we construct counterfactual representations $\tilde{z}_{cf} = g(c_{path}, s_{rcm})$ only during training as a causal consistency constraint. Aligning RCM content features with pathology-derived content encourages reliance on invariant pathological cues despite modality variation. With cross-modal alignment and weak intervention, the model learns stable micro-scale causal embeddings that reduce noise and artifacts and support reliable instance-level representations for patient-level aggregation. The final training objective combines patient-level classification with the alignment, reconstruction, content-style orthogonality, and intervention-consistency losses. Unless otherwise specified, auxiliary CMCD losses are assigned equal weights, and PMCM hyperparameters are reported in Sec. 3.1.

2.3 Patient-aware Macro-level Causal Modeling

Patient-level MIL models often capture non-causal intra-patient macro patterns, such as anatomical location, pigmentation, and chronic tissue context. These patterns exhibit intra-patient consistency but inter-patient variability, potentially inducing spurious correlations that impair cross-patient stability. Since direct predictive invariance constraints may reduce discriminability by collapsing patient-specific variations into near-constant representations, we propose PMCM, a patient-aware macro-level causal modeling strategy. Specifically, PMCM attenuates non-causal macro confounders shared within a patient while preserving discriminative components, thereby ensuring robust stability without sacrificing representational strength.

For patient p with N_p images, let the disentangled content features be $\{H_j\}_{j=1}^{N_p} \subset \mathbb{R}^D$. We model recurrent intra-patient stable factors as a low-rank invariant structural subspace. We first compute the intra-patient mean and covariance:

$$\mu_p = \frac{1}{N_p}\sum_{j=1}^{N_p} H_j, \quad \Sigma_p = \frac{1}{N_p}\sum_{j=1}^{N_p}(H_j - \mu_p)(H_j - \mu_p)^\top \tag{3}$$

We then conduct eigenvalue decomposition of the patient-specific covariance matrix Σ_p and select the Top $-$ k eigenvectors corresponding to the largest eigenvalues to construct the invariant structural subspace of patient p :

$$B_p = \text{Top} - k \text{ eigvecs}(\Sigma_p), \quad B_p \in \mathbb{R}^{D\times k} \tag{4}$$

The top-k eigenvectors of Σ_p form a k-dimensional low-rank subspace B_p that captures dominant intra-patient high-variance directions, which primarily correspond to non-causal structural factors (e.g., anatomy and pigmentation), while sparse pathology-specific signals tend to reside outside this subspace. Thus, PMCM estimates patient context directly from the covariance of instances in each bag, without introducing a separate context encoder.

We estimate instance-level invariant consistency by projecting the content embedding H_i of the i-th instance from patient p onto the invariant subspace B_p, yielding projection coefficients $c_i = B_p^\top H_i$.

The L2 norm of this coefficient vector is then used as a measure of the magnitude of variation of the instance along the patient-invariant structural directions:

$$g_i = \sigma(\text{MLP}(\|c_i\|_2)), \quad g_i \in (0, 1) \tag{5}$$

This weight quantifies variation along patient-shared invariant structural directions. Accordingly, we apply direction-sensitive modulation to the instance embedding:

$$H_i' = H_i - \lambda_g \cdot g_i \cdot B_p c_i \tag{6}$$

The modulation suppresses each instance's projection onto the patient-specific invariant subspace. As invariant macro-structural components reflect confounders, this procedure reduces patient bias while preserving pathology-related signals.

Although invariant directions are mitigated, pathology-related signals may still vary across instances. We introduce a patient-level causal budget normalization scheme for competitive reweighting to avoid domination by background-consistent instances.

Concretely, the causal contribution of instance i is measured according to the magnitude of its direction-modulated embedding:

$$\alpha_i = \frac{\|H_i'\|_2}{\sum_{j=1}^{N_p}\|H_j'\|_2} \cdot N_p \tag{7}$$

The feature amplitudes are subsequently redistributed under a fixed patient-level representational budget:

$$H_i^{\text{causal}} = \lambda \cdot \alpha_i \cdot H_i' \tag{8}$$

Table 1. Quantitative comparison of our proposed method with other representative MIL methods on the public RCM dataset. "P", "S", and "C" indicate the use of pathology guidance, structural context modeling, and causal constraints, respectively.

Method	P	S	C	AUC	ACC	F1
AttenMIL	✗	✗	✗	0.8699	0.8028	0.8018
ABMIL[15]	✗	✗	✗	0.8602	0.8310	0.8207
DSMIL[16]	✗	✗	✗	0.8504	0.8169	0.8115
TransMIL[17]	✗	✗	✗	0.8894	0.8451	0.8431
MicroMIL[18]	✗	✓	✗	0.9057	0.8451	0.8446
IRM[19]	✗	✗	✓	0.8862	0.8169	0.8163
Ours	✓	✓	✓	**0.9447**	**0.8732**	**0.8728**

Table 2. The ablation study on the public RCM dataset. "Mi" and "Ma" denote the micro-level CMCD module and the macro-level PMCM module, respectively. Best results are highlighted in bold.

Mi	Ma	AUC	ACC	F1
✗	✗	0.8699	0.8028	0.8018
✗	✓	0.8935	0.8451	0.8405
✓	✗	0.9089	0.8451	0.8419
✓	✓	**0.9447**	**0.8732**	**0.8728**

The parameter λ regulates the degree of competition. Through competitive redistribution, the mechanism suppresses patient-invariant background dominance during MIL aggregation, improving discriminability and cross-patient stability. The resulting patient-level embedding is fed into a temperature-controlled Attention-MIL aggregator with temperature τ for final prediction. This achieves macro-level causal disentanglement by restructuring the embedding space, enhancing robustness while preserving discriminative strength.

3 Experiment

3.1 Datasets

Experiments are conducted on three datasets spanning RCM and histopathology modalities. The public Monnier RCM cohort [20] (134/30/71 patient-wise split) is the main patient-level classification benchmark because valid testing requires RCM bags with patient-level positive/negative labels. The SkinCancer dataset [21] is used only for histopathological supervision (BCC vs. non-neoplastic), not as a second RCM classification test set. An independent private RCM cohort (86 BCC-positive patients) is used for no-finetuning stress testing under RCM domain shift; because it contains only BCC-positive patients, it is not a standard OOD classification benchmark. The model is trained on a NVIDIA 4070S GPU using AdamW for 50 epochs (batch size 4, learning rate 1e-5, weight decay 0.05) using mixed precision. A ConvNeXtV2-Tiny backbone pre-trained on ImageNet-1K is adopted, with all RCM and pathology tiles resized to 224×224; up to 16 RCM patches per patient form each bag, and class-matched pathology tiles provide non-registered teacher supervision for alignment. In DualCaMIL, k=5, λ_g=0.3, λ=0.5, τ=2, selected based on validation performance.

3.2 Comparative Experiment

We compare with representative MIL baselines, including attention-based methods (Attention-MIL, ABMIL, DSMIL, TransMIL) and structurally or causally

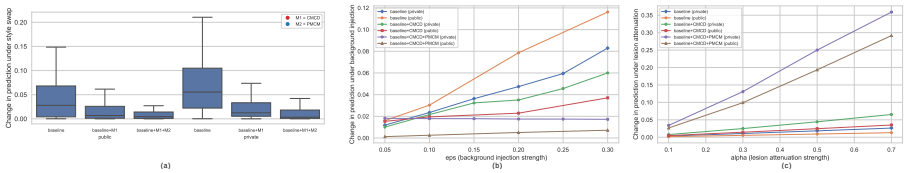

Fig. 3. PCST results on public and private RCM datasets. (a) Micro-scale imaging style invariance; (b) patient-invariant structural intervention; (c) lesion-level causal sensitivity.

motivated approaches (MicroMIL, IRM). Results are summarized in Table 1. IRM is included as a compatible invariant-learning baseline; many domain-invariant or causal alternatives require explicit environments, multi-environment training, or image/lesion-level labels unavailable in this patient-level RCM protocol. Conventional MIL models rely on instance aggregation without modeling patient-level confounders, limiting robustness in heterogeneous RCM scenarios. Structurally or causally motivated baselines (MicroMIL, IRM) partially alleviate this issue but fail to jointly address cross-modal and patient-level confounding. By integrating micro-level cross-modal disentanglement with macro-level causal modeling, DualCaMIL achieves the best performance (AUC 0.9447), demonstrating the effectiveness of unified dual-level modeling under this protocol.

3.3 Ablation Experiment

Table 2 presents ablation results for CMCD and PMCM. Removing both modules reduces AUC to 0.8699. PMCM improves robustness by modeling patient-level structure, whereas CMCD enhances lesion discrimination via cross-modal disentanglement. Their combination yields the strongest performance, demonstrating complementary effects.

3.4 Patient-level Counterfactual Stress Tests and OOD Robustness

To evaluate whether the model behaves consistently with the intended causality-inspired mechanisms at inference, we perform Patient-level Counterfactual Stress Tests (PCST) along three dimensions on the public RCM dataset.

Micro-scale imaging style invariance. We replace s_{rcm} with style features from the private dataset under different imaging conditions while preserving content. As shown in Fig. 3(a), the full model exhibits markedly smaller prediction shifts than Attention-MIL, suggesting suppression of style-induced confounders.

Patient-invariant structural intervention. We substitute the covariance-estimated patient structural subspace while keeping instance-level causal features fixed. As shown in Fig. 3(b), the full model remains stable, whereas non-patient-aware variants display larger prediction deviations, indicating mitigation of patient-level confounding.

Lesion-level causal sensitivity. We progressively attenuate high-causal-score instances. Fig. 3(c) shows rapid and monotonic prediction changes for the full model, reflecting reliance on lesion-specific evidence.

On the private RCM cohort, the full model also remains more stable under style and structural interventions while retaining lesion sensitivity, supporting robust dual-level behavior under domain shift.

4 Conclusion

We propose DualCaMIL, a dual-level causality-driven multi-instance learning framework for patient-level RCM diagnosis that addresses multi-level confounding. DualCaMIL disentangles non-causal components at both instance and patient levels. At the micro level, cross-modal guided causal disentanglement mitigates imaging-related interference, while at the macro level, patient-invariant structural factors are explicitly modeled and suppressed to prevent reliance on contextual bias. Quantitative evaluation and counterfactual stress testing demonstrate mechanism-consistent robustness to non-causal perturbations while preserving lesion-specific sensitivity. The current evidence supports patient-level RCM BCC diagnosis under the evaluated protocol rather than formal causal identification, while broader compatible patient-level MIL tasks and stronger multi-environment baselines remain important directions for future work.

Ethics Approval. The private RCM cohort was approved by the Medical Ethics Committee of Wuxi Second People's Hospital (No. Y–519, 2024). Data were anonymized, and consent was waived for this retrospective study.

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

References

1. Longo, C., Guida, S., Mirra, M., Pampena, R., Ciardo, S., Bassoli, S., Casari, A., Rongioletti, F., Spa-dafora, M., Chester, J., Kaleci, S., Lai, M., Magi, S., Mazzoni, L., Farnetani, F., Stanganelli, I., Pellacani, G.: Dermatoscopy and reflectance confocal microscopy for basal cell carcinoma diagnosis and diagnosis prediction score: A prospective and multicenter study on 1005 lesions. J. Am. Acad. Dermatol. **90**(5), 994–1001 (2024). https://doi.org/10.1016/j.jaad.2024.01.035
2. Malvehy, J., Pellacani, G.: Dermoscopy, confocal microscopy and other non-invasive tools for the diagnosis of non-melanoma skin cancers and other skin conditions. Acta Derm. Venereol. Suppl. **218**, 22–30 (2017). https://doi.org/10.2340/00015555-2720
3. Levine, A., Markowitz, O.: Introduction to reflectance confocal microscopy and its use in clinical practice. JAAD Case Rep. **4**(10), 1014–1023 (2018). https://doi.org/10.1016/j.jdcr.2018.09.019
4. Garbarino, F., Migliorati, S., Farnetani, F., De Pace, B., Ciardo, S., Manfredini, M., Reggiani Bonetti, L., Kaleci, S., Chester, J., Pellacani, G.: Nodular skin lesions: correlation of reflectance confocal microscopy and optical coherence tomography features. J. Eur. Acad. Dermatol. Venereol. **34**(1), 101–111 (2020). https://doi.org/10.1111/jdv.15953

5. Agero, A.L., Busam, K.J., Benvenuto-Andrade, C., Scope, A., Gill, M., Marghoob, A.A., González, S., Halpern, A.C.: Reflectance confocal microscopy of pigmented basal cell carcinoma. J. Am. Acad. Dermatol. **54**(4), 638–643 (2006). https://doi.org/10.1016/j.jaad.2005.11.1096
6. Cendre, R., Mansouri, A., Perrot, J.-L., Cinotti, E., Marzani, F.: Classification of lentigo maligna at patient-level by means of reflectance confocal microscopy data. Appl. Sci. **10**, 2830 (2020). https://doi.org/10.3390/app10082830
7. Rasal, R., Castro, D.C., Pawlowski, N., Glocker, B.: Deep structural causal shape models. In: Karlinsky, L., Michaeli, T., Nishino, K. (eds.) Computer Vision – ECCV 2022 Workshops. LNCS, pp. 400–432. Springer, Cham (2023)
8. Pawlowski, N., Castro, D.C., Glocker, B.: Deep structural causal models for tractable counterfactual inference. In: Proceedings of the 34th International Conference on Neural Information Processing Systems (NeurIPS 2020), articleno. 73 (2020)
9. Zhang, Z., Wang, B., Yao, L., Demir, U., Jha, D., Turkbey, I.B., Gong, B., Bagci, U.: Domain generalization with adversarial intensity attack for medical image segmentation. arXiv:2304.02720 (2023)
10. Pocevičiūtė, M., Eilertsen, G., Garvin, S., Lundström, C.: Detecting domain shift in multiple instance learning for digital pathology using Fréchet domain distance. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. LNCS, pp. 157–167. Springer, Cham (2023)
11. Wu, X., Wang, H., Wu, H.: Causal attention multiple instance learning for whole slide image classification. In: Nguyen, V., Lin, H.-T. (eds.) Proceedings of the 16th Asian Conference on Machine Learning. Proc. Mach. Learn. Res., vol. 260, pp. 984–999 (2025)
12. Nori, S., Rius-Díaz, F., Cuevas, J., Goldgeier, M., Jaen, P., Torres, A., González, S.: Sensitivity and specificity of reflectance-mode confocal microscopy for in vivo diagnosis of basal cell carcinoma: a multicenter study. J. Am. Acad. Dermatol. **51**(6), 923–930 (2004). https://doi.org/10.1016/j.jaad.2004.06.028
13. Gareau, D.S., Patel, Y.G., Li, Y., Aranda, I., Halpern, A.C., Nehal, K.S., Rajadhyaksha, M.: Confocal mosaicing microscopy in skin excisions: a demonstration of rapid surgical pathology. J. Microsc. **233**(1), 149–159 (2009). https://doi.org/10.1111/j.1365-2818.2008.03105.x
14. Gareau, D., Bar, A., Snaveley, N., Lee, K., Chen, N., Swanson, N., Simpson, E., Jacques, S.: Tri-modal confocal mosaics detect residual invasive squamous cell carcinoma in Mohs surgical excisions. J. Biomed. Opt. **17**(6), 066018 (2012). https://doi.org/10.1117/1.JBO.17.6.066018
15. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: Dy, J., Krause, A. (eds.) Proceedings of the 35th International Conference on Machine Learning. Proc. Mach. Learn. Res., vol. 80, pp. 2127–2136 (2018)
16. Li, B., Li, Y., Eliceiri, K.W.: Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 14318–14328 (2021)
17. Shao, Z., Bian, H., Chen, Y., et al.: TransMIL: transformer based correlated multiple instance learning for whole slide image classification. In: Advances in Neural Information Processing Systems (NeurIPS), vol. 34, pp. 2136–2147 (2021)
18. Kim, J.W., Wong, B., Fu, H., et al.: MicroMIL: graph-based contextual multiple instance learning for patient diagnosis using microscopy images. arXiv:2407.21604 (2024)
19. Arjovsky, M., Bottou, L., Gulrajani, I., et al.: Invariant risk minimization. arXiv:1907.02893 (2019)

20. Campanella, G., Navarrete-Dechent, C., Liopyris, K., Monnier, J., Aleissa, S., Minhas, B., Scope, A., Longo, C., Guitera, P., Pellacani, G., Kose, K., Halpern, A.C., Fuchs, T.J., Jain, M.: Deep learning for basal cell carcinoma detection for reflectance confocal microscopy. J. Invest. Dermatol. **142**(1), 97–103 (2022). https://doi.org/10.1016/j.jid.2021.06.015
21. Kriegsmann, K., Lobers, F., Zgorzelski, C., Kriegsmann, J., Janßen, C., Meliß, R.R., Muley, T., Sack, U., Steinbuss, G., Kriegsmann, M.: Deep learning for the detection of anatomical tissue structures and neoplasms of the skin on scanned histopathological tissue sections. Front. Oncol. **12**, 1022967 (2022). https://doi.org/10.3389/fonc.2022.1022967