

Causal Shortcut Reduction for Stable Multi-Center Histopathology Image Diagnosis

Authors

Sarah Thompson^{1*}, Daniel Chen², Marc Tremblay³

Affiliations

Department of Pathology and Laboratory Medicine, Faculty of Medicine, University of British Columbia, Vancouver, BC V5Z 1M9, Canada

*Corresponding author: S.Thompson@pathology.ubc.ca

Abstract

Histopathology image diagnosis is vulnerable to shortcut learning because neural networks may capture staining protocols, scanner artifacts, tissue preparation differences, or hospital-specific color distributions rather than disease-related cellular morphology. This study examines causal shortcut reduction for robust tumor classification across heterogeneous pathology centers. We propose a Causal Morphology Constraint Network (CMCN), which identifies shortcut-sensitive features through inter-center risk variation, stain perturbation response, and patch-level attribution consistency. The model then suppresses shortcut dependence using morphology-preserving counterfactual augmentation and an invariant risk penalty across acquisition environments. Experiments were conducted on 86,420 whole-slide image patches collected from 12,760 patients across five medical centers, covering benign tissue, low-grade lesions, and malignant tumors. External validation used two unseen hospitals with different staining machines and scanning resolutions. Compared with a standard DenseNet-121 model, CMCN improved external AUC from 0.781 to 0.846 and reduced the performance gap between internal and external testing from 12.8% to 5.3%. Under synthetic stain shift, classification accuracy decreased by only 3.9%, compared with 14.6% for the baseline model. These results suggest that causal shortcut reduction can improve diagnostic stability in pathology image analysis.

Keywords: Shortcut learning; causal representation; histopathology diagnosis; domain shift; tumor classification; invariant learning

1. Introduction

Histopathological examination remains a fundamental component of tumor detection, grading, prognosis assessment, and treatment planning. The increasing adoption of digital whole-slide imaging has further enabled computational models to analyze tissue architecture and cellular morphology at both patch and slide levels. Convolutional neural networks have demonstrated strong performance in extracting discriminative histological features from digitized slides [1]. Multiple-instance learning has extended this capability by allowing slide-level classification from

weak labels without requiring exhaustive annotation of every image region [2]. Self-supervised learning has also become increasingly important because large pathology archives contain substantial amounts of unlabeled data. Contrastive learning can generate transferable representations from unlabeled tissue patches [3-6]. More broadly, causally motivated studies have shown that predictive models may simultaneously rely on multiple shortcut features and that identifying and removing these shortcuts can improve robustness under distribution shifts [4]. This issue is especially relevant in computational pathology because non-biological image characteristics can be strongly correlated with diagnostic labels during training. Transformer-based pathology models, including CTransPath and hierarchical architectures, have further improved the modeling of tissue structure across spatial scales and magnifications [5,6]. Large pathology foundation models have recently extended this direction by pretraining on extensive collections of whole-slide images and demonstrating transferable performance across several downstream tasks [7,8]. Despite these advances, high accuracy on an internal test set does not necessarily indicate that a model has learned stable disease-related morphology[9-12].

Histopathology images are influenced by numerous technical and institutional factors that are unrelated to the underlying disease. Fixation protocols, section thickness, staining duration, reagent concentration, laboratory procedures, scanner manufacturer, optical settings, image compression, and spatial resolution can all alter the visual appearance of a slide. These variations may change color distributions, texture statistics, contrast, edge sharpness, and image noise without changing the biological state of the tissue. A highly flexible model can therefore exploit such acquisition-related characteristics if they correlate with diagnostic labels in the training data. In this setting, stain color, scanner noise, tissue borders, focus patterns, or preparation artifacts may act as shortcuts that provide an easier predictive signal than disease-related nuclear and tissue morphology[13-16]. Deep representations extracted from pathology images have been shown to contain sufficient information to identify the originating medical center [9]. This finding suggests that institutional characteristics can become embedded in learned representations and may artificially increase performance when training and testing data are drawn from the same hospitals. Feature-selection strategies can reduce some forms of institutional bias [10], but their effectiveness depends on whether the relevant shortcut-sensitive features are known and correctly identified. Independent evaluation across patients and hospitals is therefore essential for assessing the actual generalizability of pathology models [11,12]. The problem is particularly important in multi-center deployment because a model that performs well within one laboratory environment may deteriorate when staining protocols, scanners, or tissue-processing procedures change. Existing approaches to this problem mainly focus on stain normalization, stain augmentation, domain generalization, and shortcut detection. Stain-normalization methods attempt to map tissue images into a more consistent color space, whereas augmentation methods deliberately introduce appearance variation during training[17-20]. HistAuGAN generates alternative stain styles while preserving tissue structure [13], and RandStainNA combines stain normalization and stochastic

augmentation across multiple color spaces [14]. Automatic H&E augmentation has also been reported to improve classification performance across centers by increasing the range of staining conditions observed during training [15]. These methods can reduce sensitivity to color differences, but stain variation represents only one component of institutional shift. Scanner artifacts, focus quality, compression patterns[21-25], tissue sectioning, illumination characteristics, and laboratory-specific preparation procedures may remain unchanged after color normalization. Excessively strong color transformations may also disturb diagnostically useful nuclear or cytoplasmic information, which creates a trade-off between suppressing technical variation and preserving biological morphology[26-30]. Domain-generalization methods attempt to address a broader range of shifts by learning representations that remain effective on previously unseen hospitals or scanners [16,17]. Shortcut-testing approaches provide another strategy by evaluating whether predictions change when known nuisance variables are manipulated or removed [18,19]. These approaches have provided useful evidence that technical artifacts can influence pathology classifiers, but most studies investigate only one dominant source of variation at a time. A color-based shortcut, for instance, may coexist with scanner noise[31-34], tissue-border patterns, preparation artifacts, or center-specific imaging characteristics [20,21]. Controlling only one nuisance factor may therefore leave other shortcut relationships intact. The challenge becomes more difficult because shortcut features are not always visually obvious or statistically independent [22]. Several acquisition-related factors may occur together within the same hospital, allowing the model to construct a composite representation of institutional identity rather than relying on a single artifact[35-37]. A staining pattern may interact with scanner resolution, image sharpness, or tissue-processing characteristics and create a highly predictive but non-biological signal. Standard feature-importance methods cannot reliably distinguish such signals from genuine tumor morphology because they describe how a fitted model uses information rather than whether that information represents a stable disease mechanism. Similarly, a model may retain high internal AUC even when its attention gradually shifts away from nuclei or tissue architecture toward peripheral regions[38-40], empty background, or preparation artifacts [23,24]. Many published studies rely on public datasets collected from one institution, a small number of centers, or randomly partitioned samples. These designs can unintentionally preserve the same acquisition patterns in both training and test sets. As a result, reported accuracy may reflect similarity in laboratory conditions rather than true biological generalization. A more rigorous evaluation should therefore combine independent hospitals with controlled perturbations that alter acquisition-related appearance while preserving the underlying morphology. It should also examine whether the spatial evidence used by the model remains consistent after these perturbations rather than evaluating classification performance alone[41-43].

To address these limitations, this study proposes a Causal Morphology Constraint Network (CMCN) for stable tumor classification across pathology centers. The objective of CMCN is to reduce dependence on acquisition-related shortcuts while preserving morphological features that

remain informative across institutions. The framework evaluates shortcut sensitivity from complementary perspectives, including variation in prediction risk across medical centers, changes in feature responses under controlled stain perturbations, and the consistency of patch-level attribution across environments. Features that show large center-dependent risk variation, excessive sensitivity to acquisition-related appearance changes, or unstable spatial attribution are treated as potential shortcut features. Their influence is reduced through morphology-preserving counterfactual augmentation and an invariant risk penalty. The counterfactual augmentation modifies acquisition-related appearance while retaining nuclear organization, cellular structure, and tissue architecture, allowing the model to observe alternative image appearances without changing the underlying pathological content. The invariant risk penalty further encourages the classifier to rely on representations that remain predictive across different institutional environments. This design is informed by recent advances in causal representation learning, but adapts causal robustness specifically to the multi-source technical variation present in histopathology images. The experiments include 86,420 patches from 12,760 patients collected at five medical centers, while data from two additional hospitals are reserved exclusively for external validation. The dataset contains benign tissue, low-grade lesions, and malignant tumors and includes differences in staining equipment, laboratory procedures, and scanning resolution. Performance is evaluated using external AUC, the gap between internal and external performance, and robustness under controlled synthetic stain shifts. Changes in patch-level attribution are also examined to determine whether improved generalization is associated with more stable attention to disease-relevant morphology. The study therefore investigates whether direct identification and suppression of shortcut-sensitive features can provide more reliable tumor classification than standard empirical training or stain augmentation alone. More broadly, the work aims to establish a pathology-specific framework for evaluating model robustness in terms of both predictive performance and morphological consistency. By linking causal shortcut control with cross-center validation, CMCN provides a practical approach for developing histopathology models that are less dependent on institutional acquisition patterns and more likely to maintain stable diagnostic performance when deployed in previously unseen clinical environments.

2. Materials and Methods

2.1 Study Cohort and Setting

The retrospective dataset contained 86,420 image patches from 12,760 patients at five medical centers. Each patient had a final diagnosis of benign tissue, a low-grade lesion, or a malignant tumor. The slides came from biopsy or surgical samples and were stained with hematoxylin and eosin. The centers used different staining machines, slide scanners, and scanning resolutions. These differences were kept in the dataset to reflect hospital-level variation. Slides were included if the diagnosis was available and enough tissue was present. Slides with severe folds, blur, tissue damage, or poor staining were excluded. Patient age, sex, sample type, and

medical center were recorded but were not used as model inputs. Data from two other hospitals were reserved for external testing. Their staining and scanning settings were not included in model training.

2.2 Experimental Design and Control Models

Patients from the five development centers were assigned to training, validation, and internal test sets at a ratio of 70:15:15. The class and center distributions were balanced across the three sets. All patches from one patient were placed in the same set to prevent data leakage. The two external cohorts were not used for training or model selection. A DenseNet-121 model with standard image augmentation was used as the main control. A second control used stain augmentation without shortcut removal. Three reduced versions of CMCN were also evaluated. They removed counterfactual augmentation, the invariant risk penalty, or the attribution consistency term, respectively. All models used the same data splits, initial weights, optimizer, and stopping rule. Models were trained with AdamW using an initial learning rate of 1×10^{-4} , a batch size of 64, and no more than 100 epochs. Training stopped if the validation loss did not decrease for 10 epochs.

2.3 Image Measurement and Quality Control

Whole-slide images were processed at $20\times$ magnification. Tissue areas were separated from the white background in the HSV color space. Patches of 224×224 pixels were then taken from the detected tissue areas. Images were resized when scanner resolution differed. Patches with little tissue, severe blur, pen marks, folds, bubbles, or abnormal staining were removed using fixed quality limits. Two pathologists checked the diagnostic labels and reviewed a random sample of included and excluded patches. Disagreements were settled through joint review. The final pathology report was used as the correct diagnosis. Pixel values were scaled using the mean and standard deviation of the training set. Information from the validation and test sets was not used during this step. Counterfactual images were produced by changing hematoxylin and eosin levels within the range found in the five development centers. Nuclear boundaries and tissue structure were kept unchanged. Images that failed the structure and edge checks were removed.

2.4 Data Processing and Model Formulation

DenseNet-121 was used to extract image features. Global average pooling and a three-class output layer were used for classification. CMCN calculated a shortcut score for each feature channel. The score was defined as follows: $s_k = \alpha \text{Vare}(\Delta \text{Re}, k) + \beta \frac{1}{N} \sum_{i=1}^N |z_{i,k} - \tilde{z}_{i,k}| + \gamma [1 - \text{Corr}(A_i, \tilde{A}_i)]$. Here, s_k is the shortcut score for feature channel k . The term $\Delta \text{Re}, k$ is the change in classification loss after channel k is removed from center e . The variables $z_{i,k}$ and $\tilde{z}_{i,k}$ are the feature values of the original and stain-changed images, respectively. The variables A_i and $\tilde{A}_i$ are their attribution maps. A high value of s_k indicates that the feature may depend on the medical center or staining condition. Features with high shortcut scores were assigned lower

weights. The total training loss was defined as follows:
 $\mathcal{L} = \mathcal{L}_{CE} + \lambda_{cf} \frac{1}{N} \sum_{i=1}^N D_{KL}(p_i // p_i^{cf}) + \lambda_{inv} \frac{1}{E} \sum_{e=1}^E (R_e - \bar{R})^2$. Here, $\mathcal{L}_{CE}$ is the weighted classification loss. The term $D_{KL}(p_i // p_i^{cf})$ measures the difference between the predictions for the original and counterfactual images. The variables R_e and $\bar{R}$ denote the center-specific risk and the mean risk across all centers, respectively. The coefficients λ_{cf} and λ_{inv} control the two penalty terms. Their values were selected using the validation set.

2.5 Statistical Analysis

Patch probabilities were averaged to obtain one prediction for each patient. The main outcome was the macro-average area under the receiver operating characteristic curve. Other measures included accuracy, macro-F1 score, sensitivity, specificity, Brier score, and calibration error. The difference between internal and external performance was also calculated. Patient-level bootstrap sampling with 2,000 repeats was used to calculate 95% confidence intervals. AUC values were compared with the DeLong test. Accuracy values were compared with McNemar's test. Resistance to stain change was measured as the decrease in accuracy between original and stain-changed images. Results were examined by hospital, diagnostic class, sample type, and scanner resolution. Each experiment was repeated with five random seeds. All tests were two-sided. $P < 0.05$ was treated as statistically significant. The Holm method was used when several comparisons were performed.

3. Results and Discussion

3.1 Overall Classification Performance

CMCN performed better than the standard DenseNet-121 model. On the two external hospital datasets, the AUC increased from 0.781 to 0.846. The absolute increase was 0.065. Both hospitals used staining machines and scanning resolutions that were not included in the training data. Therefore, the performance gain was maintained outside the five training centers. As shown in Fig. 1, DenseNet-121 had a larger performance loss after transfer to the external hospitals. The baseline model may have learned color patterns, scanner noise, or tissue preparation features that were linked to certain centers. CMCN reduced the effect of these features by giving lower weights to shortcut-sensitive channels. The model therefore placed more weight on cell shape and tissue structure. A similar problem was reported by Emam et al. (2026). Their study found that site-related image features could increase internal accuracy but reduce accuracy when data from different centers were tested separately [25,26].

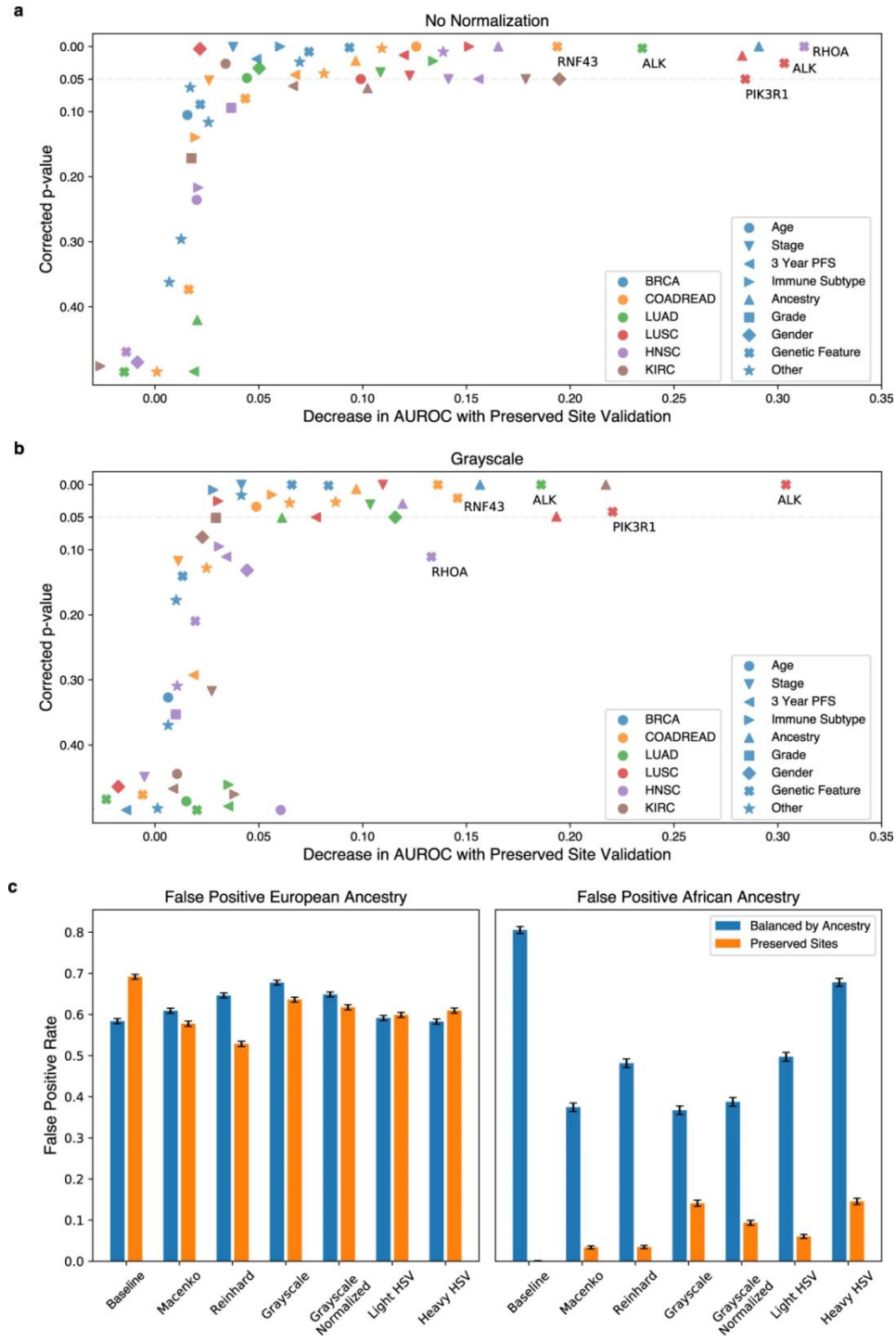


Fig. 1. External AUC and internal-external performance gap for DenseNet-121 and CMCN.

3.2 Stability across Medical Centers

The performance gap between internal and external testing decreased from 12.8% for DenseNet-121 to 5.3% for CMCN. This was a decrease of 7.5 percentage points, or 58.6% relative to the baseline. A smaller gap means that the model produced more stable results across hospitals. It also shows that internal testing alone may overestimate model performance. Bilal et al. (2026) reported an average AUROC decrease of 0.069 when standard cross-validation was replaced with

center-separated validation. They also found lower performance for 51 of 56 predictable clinical or molecular outcomes. The DenseNet-121 results in this study followed the same pattern. CMCN reduced this loss through inter-center risk control and feature weighting. However, the remaining 5.3% gap shows that center effects were not fully removed. Differences in patient groups, tissue processing, and scanner settings may still affect the model [27,28].

3.3 Response to Stain Changes

Under synthetic stain changes, the accuracy of DenseNet-121 decreased by 14.6%, whereas the accuracy of CMCN decreased by only 3.9% (Fig. 2). The difference was 10.7 percentage points. This represents a 73.3% reduction in accuracy loss compared with the baseline. The result supports the use of morphology-preserving counterfactual images [29,30]. These images changed the stain color but kept the cell borders, nuclear shape, and tissue structure unchanged. As a result, color became less useful as a shortcut during training. The attribution consistency term also kept the model focused on similar tissue regions before and after the stain change. Ceylan et al. (2026) also reported better cross-center classification using Fourier-based domain adaptation than using color normalization or standard augmentation. Their method aligned source and target features. In contrast, CMCN identified and controlled shortcut-sensitive channels without using data from the two external hospitals.

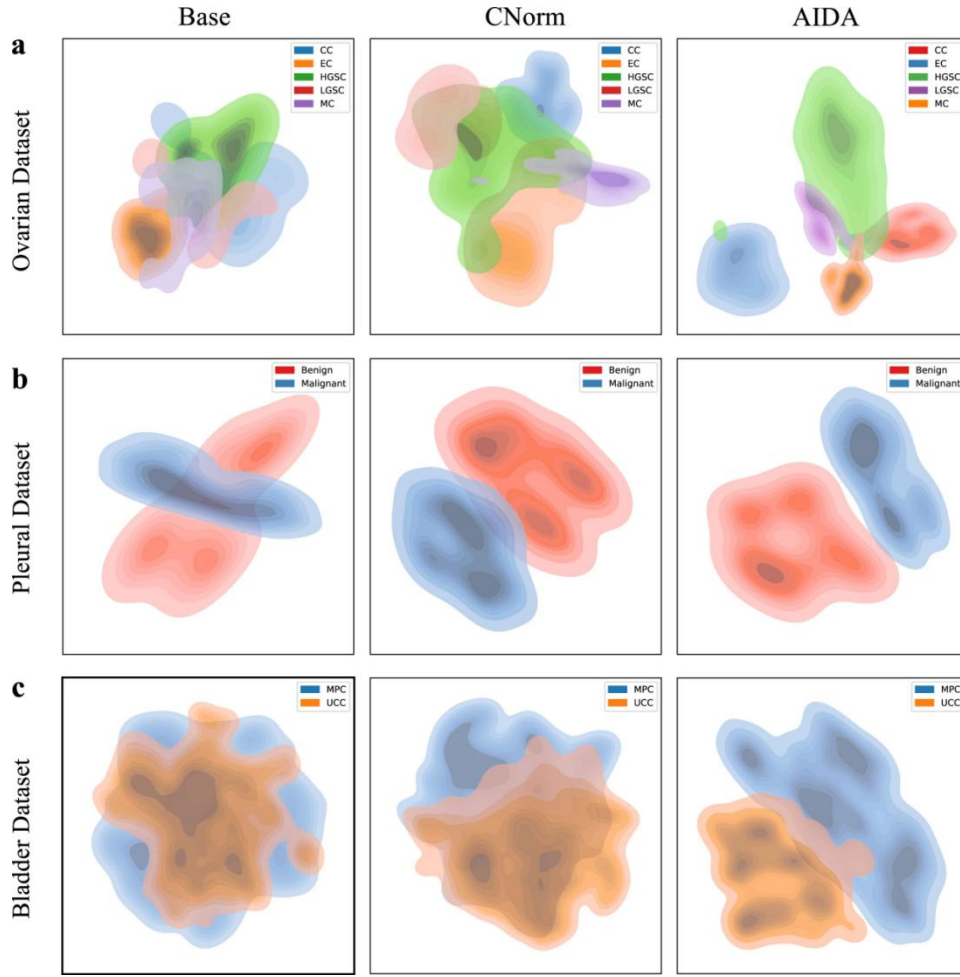


Fig. 2. Accuracy decrease of DenseNet-121 and CMCN after synthetic stain changes.

3.4 Clinical Meaning and Study Limitations

The results indicate that shortcut control can improve tumor classification across different pathology centers. The study included 86,420 patches from 12,760 patients. Data from five centers were used for model development, and data from two other hospitals were used for external testing. This design tested both real hospital differences and synthetic stain changes. However, the study has several limitations. First, the data were collected retrospectively. The seven hospitals may not represent all staining systems and scanners used in clinical practice. Second, synthetic stain changes cannot reproduce tissue folds, focus errors, image compression, or poor tissue fixation. Third, the three diagnostic classes were broad. More testing is needed for tumor types that have similar tissue features. Finally, reducing measured shortcut use does not prove that every feature retained by the model is directly related to disease. Future studies should use prospective data and more medical centers. They should also examine patient-level accuracy, probability calibration, subgroup results, and common failure cases.

4. Conclusion

CMCN was developed to reduce shortcut learning in multi-center histopathology classification. The method measured feature responses to center differences, stain changes, and attribution maps. Features linked to these unwanted image patterns received lower weights. Counterfactual images and a risk penalty were also used during training. On the external datasets, CMCN increased the AUC from 0.781 to 0.846. It also reduced the internal–external performance gap from 12.8% to 5.3%. After synthetic stain changes, accuracy decreased by 3.9% for CMCN and by 14.6% for DenseNet-121. CMCN was therefore less affected by staining and hospital-specific image features. It could help provide stable tumor classification in hospitals that use different staining and scanning systems. However, this retrospective study included seven hospitals, three broad classes, and patch-level images. Synthetic changes also cannot cover all errors found in clinical slides. Future studies should test CMCN on whole-slide images, more tumor types, and prospective multi-center data.

References

1. Zhang, Z. (2026). A Study on the Impact of Cost Presentation on Decision-making Bias and Cancellation Behavior in Online Subscriptions. Available at SSRN 7349839.
2. Jalil, M. A., & Zafra, A. (2025). Deep Learning in Multiple Instance Learning: Methods, Applications and Research Trends. IEEE Access.
3. Du, Y., Chen, Z., Liu, Q., & Dong, Y. (2026). Co-identification of Vibration, Structural Looseness, and Surface Damage in Industrial Equipment Operation Videos.
4. Zheng, J., & Makar, M. (2022). Causally motivated multi-shortcut identification and removal. *Advances in Neural Information Processing Systems*, 35, 12800-12812.
5. Mini Han Wang , " AI-Powered Innovations in Ophthalmic Diagnosis and Treatment ", Bentham Science Publishers (2025). <https://doi.org/10.2174/97988988122871250101>.
6. Xiong, W., Zeng, Y., & Guo, Y. (2026). A Study on the Characterization of Building MEP System States and the Learning of Behavioral Patterns Driven by Large-Scale Operational Data. Available at SSRN 7121883.
7. Hong, Z., Liang, S., Xu, T., & Chen, H. (2026). A Study on Failover Verification and Recovery Objective Prediction for Cross-Region Cloud Services.
8. Wu, J., Zhang, J., Wu, D., & Peng, Y. (2026). Visual Transformation Mechanisms for Integrating Digital Cultural Heritage Resources into Junior High School Art Curricula.
9. Han, Z., Chen, W., Han, Y., Mao, R., & Qin, J. (2026). Fast Diversified Top-k Rule Discovery via User-Guided Embeddings. *IEEE Transactions on Knowledge and Data Engineering*.
10. Huang, J., Xu, T., Yang, J., & Yin, J. (2026). A Graph Neural Network Approach for Financial Data Validation and Risk Identification in Large-Scale Intercompany Transactions. Available at SSRN 7175058.

11. Zhang, Z., Wang, J., Li, Z., Wang, Y., & Zheng, J. (2025). Annocoder: A mti-agent-based code generation and optimization model. *Symmetry*, 17(7), 1087.
12. Gu, X. (2026). Identifying Causal Effects and Analyzing Heterogeneity of User Growth Interventions on Digital Platforms: Evidence from Large-Scale Behavioral Data. Available at SSRN 6809181.
13. Quan, Q., Gao, Y., & Wang, Q. (2025). The impact of teacher emotional support on students' engagement in AI-mediated English learning environments: The mediating role of resilience and self-efficacy. *Acta Psychologica*, 260, 105766.
14. Zhao, J., Fan, J., & Li, L. (2026). A Study on an Explainable Causal-Enhanced LLM Agent for Predicting the Forming Quality of Automotive Component Materials.
15. You, S. (2026). Verifiable Audit Mechanisms in AI Compliance Automation: Scalability. Available at SSRN 6547458.
16. Jiao, Y., Shi, T., Zhao, B., & Wang, A. (2026). Retrieval-Guided Structured Reasoning and Interpretable Representation Learning for Large-Scale Video–Language Models.
17. Wang, Z., Yu, J., Liu, H., Zheng, Z., Jin, Y., Li, S., ... & Zhang, K. (2025, July). Generative Music Models' Alignment with Professional and Amateur Users' Expectations. In *Findings of the Association for Computational Linguistics: ACL 2025* (pp. 6909-6920).
18. Liang, S., Du, Y., Chen, W., & Liu, Z. (2026). Order Allocation and Emergency Transportation Decisions Amid Multi-Source Procurement Disruptions.
19. Bao, Y., & Wang, H. (2026). Orion: A High-Throughput, Fault-Tolerant Event Routing Architecture for Hyperscale Data Streams. *Fault-Tolerant Event Routing Architecture for Hyperscale Data Streams* (January 01, 2026).
20. Wang, S., Feng, Y., & Fang, X. (2026, May). A Large Language Model-Enabled Multi-Agent Collaboration Method for Complex Task Solving. In *2026 6th International Symposium on Computer Technology and Information Science (ISCTIS)* (pp. 253-256). IEEE.
21. Xu, Y., Sun, Y., Zhai, B., Li, M., Liang, W., Li, Y., & Du, S. (2025, April). Zero-shot video moment retrieval via off-the-shelf multimodal large language models. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 39, No. 9, pp. 8978-8986).
22. Yang, Q., & Du, Y. (2026). Predicting Rollback Risks and Controlling Gradual Rollout Traffic for Online Ranking Feature Deployments.
23. Zhao, Z., & Welsch, R. E. (2024). Hierarchical reinforced trader (hrt): A bi-level approach for optimizing stock selection and execution. *arXiv preprint arXiv:2410.14927*.
24. Zheng, X., Chen, X., Gong, S., Griffin, X., & Slabaugh, G. (2025, September). Xfmamba: Cross-fusion mamba for multi-view medical image classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 672-682). Cham: Springer Nature Switzerland.
25. Liang, S., Hsu, K. H., & Liu, Z. (2026). Rolling Scheduling and Capacity Balancing for Server Assembly Under Engineering Change Disturbances.

26. Yuan, Y., Xu, T., Yin, J., & Huang, J. (2026). Modeling Loading Anomalies and Identifying Root Causes in Media Consumption Workflows on Large Social Media Platforms. Available at SSRN 7393318.
27. Su, J., Cai, C., Zhu, F., He, C., Xu, X., Guan, D., & Si, C. (2024, September). Momentum auxiliary network for supervised local learning. In European Conference on Computer Vision (pp. 276-292). Cham: Springer Nature Switzerland.
28. Bao, Y., & Wang, H. (2026). Large-scale Metrics Migration: A Systematic Approach to Rebuilding Production Measurement Infrastructure. Available at SSRN 7185660.
29. Ceylan, M., Tabin, A., Langer, P., Fleisch, E., & Barata, F. (2026). EvoMorph: Counterfactual Explanations for Continuous Time-Series Extrinsic Regression Applied to Photoplethysmography. arXiv preprint arXiv:2601.10356.
30. Yang, Z., Alexandrova, A. N., & Sautet, P. (2026). Modeling CO₂ Hydrogenation to Methanol on an Ensemble of Inverse ZrO₂ on Cu Catalytic Sites: Mechanism, Reactivity, and Deactivation. *Angewandte Chemie*, e5448247.
31. Lyu, Y., Armas-Cervantes, A., & Mendoza, A. (2025, December). Uncovering self-regulated learning patterns in Chatbot-powered environment through a process mining lens. In 36th Australasian Association for Engineering Education Annual Conference (AAEE2025) (pp. 371-379). Brisbane, QLD: Engineers Australia.
32. Liang, J., Xiao, Y., Yang, H., Yu, Z., Liu, J. N., & Yang, K. (2026, April). C-HyPOD: Causal Hyperbolic Representation Learning with Prototype Orthogonal Disentanglement for Graph Out-of-Distribution Recommendation. In Proceedings of the ACM Web Conference 2026 (pp. 6541-6550).
33. Xi, X., Chen, P., Qiu, Y., Luo, R., Tong, P., & Liang, J. Video-BCI: Bayesian Cognitive Integration of Self-Prior Hypotheses for Video Understanding. In Forty-third International Conference on Machine Learning.
34. Xu, Z., Zhang, X., Li, R., Tang, Z., Huang, Q., & Zhang, J. (2024). Fakeshield: Explainable image forgery detection and localization via multi-modal large language models. arXiv preprint arXiv:2410.02761.
35. Jia, Z., Huang, C., Wang, Z., Fei, H., Wu, S., & Feng, J. (2024). Finger recovery transformer: toward better incomplete fingerprint identification. *IEEE Transactions on Information Forensics and Security*, 19, 8860-8874.
36. Huang, C., Jia, Z., Fei, H., Zhu, Y., Yuan, Z., Zhang, J., & Zhou, J. (2025, October). ArtFRD: A Fisher-Rao Mixture Metric for Generative Model Aesthetic Evaluation. In Proceedings of the 33rd ACM International Conference on Multimedia (pp. 6654-6662).
37. Cui, J., Zhang, G., Chen, Z., & Yu, N. (2022). Multi-homed abnormal behavior detection algorithm based on fuzzy particle swarm cluster in user and entity behavior analytics. *Scientific Reports*, 12(1), 22349.

38. Jiang, C. (2026). Information doesn't work? Testing experiential drivers of VR application in heritage tourism by affective-cognitive pathways. *Journal of Hospitality and Tourism Technology*, 1-28.
39. Wang, T., & Xia, Z. (2025). Stability of In-Context Learning: A Spectral Coverage Perspective. *arXiv preprint arXiv:2509.20677*.
40. Chaoyong Ma, Nan Si, Xuewei Song, Chen Liang and Haoming Luo, A Bearing Fault Diagnosis Method Based on Concise Empirical Wavelet Transform, *International Journal of Comprehensive Engineering*, 2023, 12(1), 1-6
41. Yuyan Li, Lin Chen, Mengyu Yang, Qiaomei Li and Taifu Li, Polarization Converters Composed of Metasurfaces with Independent Manipulation of Polarization and Phase, *International Journal of Comprehensive Engineering*, 2023, 12(1), 7-14
42. Jiaxun Du, Zhenbao Fu, Jiachen Liu, Huaqing Wang, Zhentao Ge and Liuyang Song, Bearing Fault Diagnosis Method Based on Attention Residual Network Under Small Sample Condition, *International Journal of Comprehensive Engineering*, 2023, 12(1), 15-24