SlideGuard: WSI Artifact Detection Benchmark

Gabriela Kaczmarek[1,2], Zuzanna Krawczyk-Borysiak[1,2], Mateusz Miller[1,3], Adam Krawczyk[4,5], Malgorzata Sokol[1], Martyna Przybylska[1], Lukasz Szymanski[6], Tomasz Markiewicz[7,2], and Zaneta Swiderska-Chadaj[1,2]

[1] IDEAS Research Institute, Warsaw, Poland
[2] Warsaw University of Technology, Warsaw, Poland
[3] Maria Sklodowska-Curie National Research Institute of Oncology, Warsaw, Poland
[4] Poznan University of Technology, Poznań, Poland
[5] AKCES NCBR (National Centre for Research and Development), Warsaw, Poland
[6] Polish Academy of Sciences, Warsaw, Poland
[7] Military Institute of Medicine, Warsaw, Poland
gabriela.kaczmarek@ideas.edu.pl

Abstract. To ensure the highest quality in cancer research and computer aided diagnostics in computational pathology, Whole Slide Image (WSI) quality should be carefully monitored. While several open-source Quality Control (QC) tools exist, their comparative performance remains largely unaudited. In this work, we present the first independent, cross-species benchmark for WSI QC solutions focused on the critical task of automated artifact detection. Recognizing that model generalization is often hindered by narrow training data, we curated a diverse artifact-centric dataset featuring 11 distinct artifact types across 6 tissue types, 2 species, and 4 independent data sources. We evaluate both pixel-level segmentation and patch-level classification methods, normalizing their outputs and comparing them to expert-annotated WSIs. Performance is quantified using Dice score, accuracy, precision, and recall within tissue containing region. Our results reveal a substantial performance gap; average artifact detection Dice scores do not exceed 0.6, with significant performance variability across dataset subcategories. To support ongoing development, we release our dataset at https://doi.org/10.5281/zenodo.20339462 and our benchmarking tool at https://www.codabench.org/competitions/16457.

Keywords: Digital Pathology · Quality Control · Artifact Detection

1 Introduction

Recent years have seen a clear shift towards digitizing the traditional histological pipeline. While many laboratories still use microscopes, Whole Slide Images (WSIs) are rapidly gaining adoption. This digital shift brings numerous benefits, including increased work efficiency, remote slide examination, easier case consultation, and simplified comparison with archived cases [3, 4]. Crucially, WSI also paves the way for algorithms to actively support the diagnostic process.

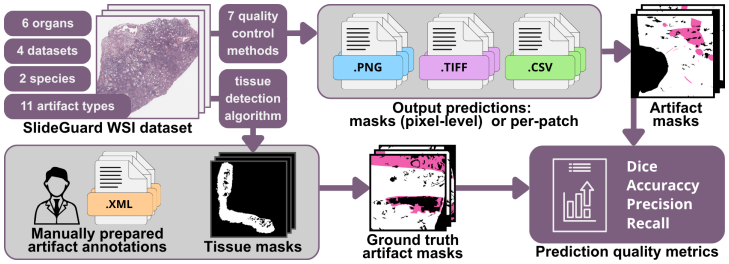

Fig. 1. Overview of the benchmarking process presented in this work.

Creating a histopathological slide involves many intricate steps, and unfortunately, each one can introduce flaws that carry over into its digital version [1, 20]. From the very beginning, after the tissue is surgically removed, it is exposed to many factors that can affect final WSI quality. During tissue processing, dehydration issues can cause cracks or empty spaces. Later, improper embedding can cause crush artifacts, altering the tissue's structure. As the paraffin block is sectioned, tissue may shred, fold, wrinkle or exhibit knife marks and chatter. Finally, even the staining process can go wrong, leaving unevenly colored areas, unwanted precipitates, or tiny air bubbles trapped under the cover-slip. When these physical slides are converted into digital images, the scanner and its software introduce their own set of problems. Common digital issues include blurry or out-of-focus regions, or noticeable lines where image patches are stitched together [11]. All artifacts, whether from physical preparation or digitization, confound the subsequent analysis.

Artifacts in WSIs do not just make it harder for experts to diagnose, slowing them down or even making a clear diagnosis impossible, but also impair the performance of automated algorithms [15, 21, 25]. If we want to successfully move AI tools from the lab into actual diagnostic use, then robust Quality Control (QC) is essential. Introducing QC as an early step in the process, especially when using ML methods, could improve reliability of the results.

Over time, many different methods have been developed to tackle the problem of artifacts. Earlier approaches often focused on detecting just one specific type of artifact, usually by using traditional image processing or carefully crafted feature extraction. For instance, Kothari et al. [10] found tissue folds by looking at color properties and using adaptive thresholds. Similarly, Wu et al. [24] pinpointed blurry areas by analyzing local pixel blur metrics with classifiers. The widely recognized HistoQC tool [6] expanded on this by combining various image metrics, handcrafted features, and supervised classifiers to identify artifact-free regions more broadly. However, since 2021, the world of QC and artifact detection has increasingly embraced Deep Learning (DL), with most new methods now built upon DL architectures [5–9, 13, 14, 16–18, 23].

Despite many new tools for checking WSI quality, we still lack good, independent benchmarks that truly test how well QC methods perform when faced with varied data. This makes it hard to build and reliably use artifact detection

tools for practical applications in pathology. To address this, we are introducing SlideGuard: the first independent, cross-species benchmark dataset that combines multiple organs, species, and data sources, specifically built to evaluate WSI artifact detection. To the authors' best knowledge, this is the first paper to provide an evaluation of recent artifact methods on a common dataset with a broad range of artifact and tissue types. Overall benchmark pipeline is illustrated on Figure 1. Our main contributions are:

- SlideGuard, a multi-source, multi-species dataset encompassing eleven artifact categories across six tissue types, designed to evaluate cross-domain generalization and model robustness.
- Thorough comparison of seven dedicated QC methods, showing their performance when faced with varied, real-world data.
- Standardized evaluation framework, offering crucial insights into current limitations and guiding the community toward developing truly robust and deployable QC solutions for the field.

2 Materials and Methods

2.1 Dataset

SlideGuard is an artifact-oriented multi-species, multi-organ, and multi-center dataset comprising 53 H&E-stained WSIs. Sourced from previous projects in accordance with data reuse policies, it includes 12 mouse WSIs from a single center [19] and 41 human WSIs sourced from three distinct centers: 25 from TCGA [2], 10 from DHMC [22], and 6 from Maria Sklodowska-Curie National Research Institute of Oncology (MSCNRIO). This enables systematic artifact benchmarking across five human tissue types (colon, prostate, lung, breast, lymphoma) and mouse skin tissue. Representative examples of these tissue types and artifact classes are shown in Figure 2.

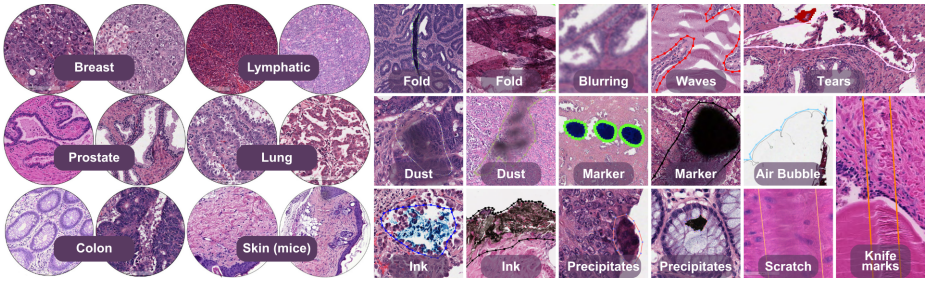

Fig. 2. Examples from the SlideGuard dataset, showing diverse tissue types (left) and various artifact classes (right).

Annotations. Using the ASAP v2.2 tool [12], an experienced histotechnologist prepared 2,116 detailed pixel level annotations for the eleven artifact classes

illustrated in Fig. 2: precipitates, tissue folds, knife marks, knife waves, scratches, ink, markers, dust, tears, air bubbles, and blurring. Additionally, SlideGuard includes 6 "High Quality" (HQ) slides selected for their near artifact-free content, i.e. exhibiting less than 0.02% artifact coverage within the tissue area.

2.2 Methods

This study benchmarks artifact detection methods for WSI quality evaluation by comparing their performance against the reference standard of pixel-level masks. This design enables a thorough investigation of method performance across diverse morphology and varying staining protocols (see Figure 2). Such structural variation makes consistent quality assessment a significant challenge. To ensure an impartial evaluation, no co-authors of this study were involved in the development of the benchmarked methods.

Compared methods selection. Benchmark inclusion was limited to methods providing either pixel-level artifact masks or patch-level classifications. We excluded approaches lacking publicly available code or model weights, as well as those providing only a single global score per WSI. This selection resulted in seven methods developed between 2019 and 2024, predominantly utilizing deep learning: GrandQC [23] (evaluated at 5x and 10x magnification), HistoQC [6], DCNN MoE [8], PathProfiler [5], a Semantic Segmentation model [18], TCNN [17], and WSIsegQC [14]. Their characteristics are compared in Table 1.

Table 1. Comparison of WSI QC methods characteristics, where: NR: Not Reported; Pred.: patch-level predictions; Src.: Sources; >X: more than X.

Method (Year)	Output	Backbone	Dataset Characteristics			
			Train Set (WSIs)	Src.	Scanners	Organs
HistoQC (2019) [6]	Mask	Gaussian NB and Random Forest	NR	NR	NR	NR
Sem. Seg.(2021) [18]	Mask	DeepLabV3+	≈ 99	NR	7	9
PathProfiler (2022) [5]	Pred.	UNET, ResNet18	107	1	1	1
TCNN (2023) [17]	Pred.	Spatial Transformer, Efficientnet B1, CNN layers	120	1	NR	22
GrandQC (2024) [23]	Mask	EfficientNetB0	420	>3	5	>7
MoE DCNN (2024) [8]	Mask	MobileNet DCNN Mixture of Experts	35	1	1	1
WSIsegQC (2024) [14]	Mask	UNet variants	350	NR	NR	>2

Workflow. The evaluation pipeline consisted of three primary stages: ground truth preparation, model inference, and performance quantification.

To prepare the ground truth, masks delineating all artifact types were generated from manual annotations. To restrict evaluation to the tissue region, these

were intersected with a tissue mask segmented via intensity thresholding at 200 (on an 8-bit scale) and refined through morphological opening and closing. After manual verification, this mask defined the Region of Interest (ROI) for all subsequent metrics, ensuring that performance quantification focused solely on artifact detection within tissue, rather than the accuracy of the methods' internal tissue-detection sub-modules.

All methods were run on all WSIs strictly using their default or recommended operational settings, without any additional parameter tuning or adjustment for the SlideGuard dataset. To assess the performance of the patch-level methods without penalizing the absence of detailed segmentation masks, in those cases the ground truth was discretized to each method's native grid; a patch was labeled as an artifact if annotation coverage exceeded 25%. For methods providing probabilistic outputs, a classification threshold of 0.5 was applied. Because the patch-level task is technically reduced to classification and inherently simplified by the discretization threshold filtering out subtle artifacts, a direct cross-mode comparison between pixel- and patch-level performance scores is limited.

Artifact detection performance was quantified using Dice scores (which mathematically reduce to F1 scores for patch-level methods), accuracy, precision, and recall. We report Artifact Dice (positive class) and Artifact-Free Dice (negative class), with the latter defined as the Dice score of the inverse artifact mask restricted to the ROI.

3 Results

Evaluation revealed significant performance variability, underscoring the complexity of automated QC across diverse WSIs. Average Dice score for each method, stratified by species and source, are detailed in Table 2. Further performance insights across sources and tissue types are illustrated in Figure 3.

Table 2. Per-category average Dice scores for QC methods, where: best results for each mode were bolded; HQ: High Quality subset with <0.02 artifact coverage; **: for this cohort, Artifact-Free Dice is reported as artifact-level metrics are uninformative in near-pristine slides.

		Pixel-level						**Patch-level**	
	Category	Histo QC	Sem. Seg.	GrandQC (5x)	GrandQC (10x)	MoE DCNN	WSIseg QC	Path Profiler	TCNN
Species	Human	0.1367	**0.3706**	0.3382	0.3351	0.1672	0.2411	0.1057	**0.2741**
	Mice	0.1766	0.2037	**0.2566**	0.2445	0.2267	0.2230	0.1538	**0.2015**
Dataset	TCGA	0.1275	**0.4206**	0.3357	0.3536	0.1393	0.2004	0.0928	**0.2259**
	MSCNRIO	0.1992	0.3844	**0.5271**	0.3261	0.3102	0.2086	0.2030	**0.3395**
	DHMC	0.1007	0.0370	0.0703	0.2326	0.1272	**0.5440**	0.0405	**0.4772**
	Mice	0.1766	0.2037	**0.2566**	0.2445	0.2267	0.2230	0.1538	**0.2015**
HQ **	DHMC	0.9283	0.7716	0.9965	0.9974	0.9813	**0.9982**	0.5805	**0.9960**

Fig. 3. Results summary. A: Radar plots showing the average Artifact Dice score (left) and average Artifact-Free Dice score (right) for the dataset (excluding HQ slides). Results are stratified by both source dataset and tissue type. B: Radar plot showing average Artifact-Free Dice and plain prediction accuracy for the HQ slides only, as artifact-level metrics are uninformative in near-pristine slides. C: Distributions of precision and recall for the dataset (excluding HQ slides), illustrating the performance across four data sources.

No single method consistently outperformed others; instead, we observed distinct strengths and weaknesses across different metrics and data divisions. While artifact-free tissue was generally segmented with high reliability, no method's average Dice score for artifact detection exceeded 0.6. This highlights a persistent difficulty in localizing artifacts precisely. Among human tissue types, GrandQC x5 [23] and the Semantic Segmentation model [18] frequently yielded the highest scores in pixel-level mode, whereas MoE DCNN [8] and HistoQC [6] exhibited lower efficacy. In patch-level mode, TCNN [17] consistently outperformed PathProfiler [5]. Performance was notably higher across all methods for prostate tissue, potentially due to its architectural homogeneity. Interestingly, in the DHMC subset, WSIsegQC [14] outperformed all alternative methods across both standard and HQ slides.

Due to divergent artifact taxonomies among the benchmarked methods, artifact type-specific precision and Dice score calculations were infeasible. Consequently, we utilized label-agnostic recall to evaluate the detection rates across our eleven artifact types. This metric assesses the proportion of artifacts correctly flagged as anomalous, regardless of the method's specific label assignment. High

recall for artifact types not included in a model's native label set indicates either cross-type false positives or a generalized sensitivity to morphological deviations from artifact-free tissue.

Table 3. Recall per artifact type for evaluated methods, where: best results for each mode were bolded

Artifact	Pixel-level						Patch-level	
	Histo QC	Sem. Seg.	GrandQC (5x)	GrandQC (10x)	MoE DCNN	WSIseg QC	Path Profiler	TCNN
Precipitates	0.3727	0.3768	0.2863	0.3565	**0.3836**	0.3338	**0.9583**	0.6939
Folds	0.3500	**0.6354**	0.5321	0.5614	0.3885	0.5235	0.7447	**0.8941**
Ink	**0.4494**	0.4197	0.0221	0.0796	0.1749	0.0167	**0.8535**	0.7051
Dust	0.3141	**0.8144**	0.5316	0.7011	0.3283	0.2193	**0.7500**	0.7308
Marker	0.7158	0.3760	0.6157	0.7315	0.5657	**0.8589**	0.9015	**0.9263**
Tears	**0.5595**	0.0995	0.1710	0.1799	0.3669	0.1724	**0.5897**	0.2515
Waves	0.0019	0.0000	0.0708	0.1141	**0.5934**	0.0300	0.0000	0.0000
Blurring	0.6157	0.3455	0.4305	0.5643	**0.7152**	0.3589	**0.9096**	0.1866
Knife marks	0.1363	0.0557	0.0111	0.0845	0.1494	**0.2086**	**0.4091**	0.0652
Scratch	**0.9728**	0.1632	0.2368	0.4154	0.4532	0.4421	**0.5641**	0.1613
Air bubble	**0.6664**	0.1555	0.3029	0.4058	0.2465	0.1794	**1.0000**	0.8899

As detailed in Table 3, among the pixel-level methods, MoE DCNN achieves the highest recall for precipitates, waves and blurring; Semantic Segmentation leads in folds and dust; HistoQC outperforms others in ink, tears, scratches, and air bubbles; and WSIsegQC yields the top scores for markers and knife marks. Within the patch-level category, PathProfiler outperforms TCNN across the majority of artifact types, with the exception of folds and markers.

4 Discussion

The integration of AI into clinical pathology necessitates robust WSI QC, as it is critical to ensuring technical standards for analysis, preventing diagnostic errors, and enabling safe deployment of automated pathology workflows. Our SlideGuard benchmark provides an evaluation of the current state of WSI artifact detection, uncovering significant challenges that demand improved solutions.

Our findings consistently reveal a substantial generalization gap in existing artifact detection methods. WSI quality assessment is inherently complex, influenced by numerous factors such as tissue staining protocols, scanner variations, slide preparation methods, and diverse tissue morphologies. These elements create significant "domain shifts" between datasets, causing models trained on limited or homogeneous data to perform poorly on unseen and varied data.

The Detection Asymmetry. A consistent discrepancy exists between a method's ability to preserve clean tissue and its precision in localizing artifacts. While high Artifact-Free Dice scores demonstrate the models' capability to preserve diagnostic tissue, the lower Artifact Dice reflects spatial boundary ambigu-

ity. This is further compounded by extreme class imbalance and the geometric imprecision of manual annotations, especially for soft-edged artifacts or subtle blur where human and machine perceptions diverge. Ultimately, neither metric is a perfect indicator of efficacy: artifact-free scores are inflated by clean tissue dominance, while artifact scores are inherently capped by the spatial limits of manual labeling regardless of model robustness.

The Generalization Gap. No method emerged as superior across all categories. While GrandQC often performed well, particularly on datasets like TCGA Prostate [2] and MSCNRIO, other methods, such as the Semantic Segmentation model or WSIsegQC, demonstrated superior performance in specific contexts (e.g., Sem. Seg. on TCGA Breast [2], WSIsegQC on DHMC [22]). This variability highlights that current methods are highly sensitive to dataset properties, such as tissue type, species, or imaging center, underscoring a lack of cross-domain robustness essential for clinical translation. Our analysis identified several factors that severely impact performance:

- **Tissue-Specific Biases:** Tissues with intricate morphologies such as lung proved more challenging than denser tissues like prostate, possibly causing models to misinterpret complex structures as artifacts.
- **Species-Level Generalization:** Interestingly, methods trained on human tissue often generalized moderately well to mouse skin tissue (e.g., GrandQC achieved Dice $\sim$0.20), performing on par with some human tissue types. This suggests that certain artifact features, such as tissue folds or tears, are morphological constants that are relatively species-agnostic. However, unique tissue characteristics and preparation differences still pose challenges for complete cross-species portability.
- **Center-Dependent Effects:** Performance showed a strong dependence on the imaging center, e.g. TCGA [2] subsets generally outperformed those from DHMC [22]. These disparities are likely caused by variations in scanner models, staining protocols, and the prevalence of specific artifact types at each center, reinforcing that models validated on a single source are likely to fail in other environments.

Towards Clinically Usable QC Tools: Practical Considerations. Beyond performance, clinical viability depends on standardizing metrics and workflow integration. A primary requirement is the shift from nominal magnification factors (e.g., 20x) to physical microns per pixel (MPP) in QC methods implementations. Because "20x" varies between 0.25 and 0.50 MPP across manufacturers, relying on fixed scaling precludes accurate downsampling and results in inconsistent morphological representations. Furthermore, interoperability remains a hurdle; many frameworks lack support for diverse formats like MRXS or rely on pre-extracted patches rather than end-to-end WSI processing. Broad adoption requires native WSI format support and seamless, end-to-end workflows.

The study acknowledges several limitations, including the dataset size and its uneven distribution across species, tissues, and centers. Residual subjectivity in artifact annotations and the occasional absence of perfectly matched tissue types for all cross-species comparisons also present inherent constraints.

5 Conclusion

This study introduces **SlideGuard**, a multi-center benchmark for WSI artifact detection. Our evaluation reveals three primary findings: first, current methods reliably identify artifact-free tissue but fail at precise artifact localization due to extreme class imbalance and annotation subjectivity. Second, model performance is inconsistent across data sources and tissue types, signaling a lack of cross-domain generalization capability. Third, reliance on nominal magnification rather than physical metrics creates morphological inconsistencies that impede usability. By providing a transparent evaluation framework, SlideGuard supports the development of next-generation, robust QC solutions for digital pathology.

Acknowledgments. We would like to thank Slawomir Pakulo, M.D., for providing the data used in this study. We gratefully acknowledge Poland's high-performance Infrastructure PLGrid (ACC Cyfronet AGH) for providing computer facilities and support within computational grants no PLG/2025/018670 and PLG/2026/019496. We also thank Marina D'Amato and prof. Francesco Ciompi for their support in running the Semantic Segmentation method to provide the results included in this work.

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

References

1. Al-Kinani, M.J.H.: Common Artifacts and Remedies in Histological Preparations. Advances in Bioscience and Biotechnology **15**(3), 174–183 (2024). https://doi.org/10.4236/abb.2024.153012
2. Cancer Genome Atlas Research Network, Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., Stuart, J.M.: The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics **45**(10), 1113–1120 (2013). https://doi.org/10.1038/ng.2764
3. Eloy, C., Fraggetta, F., van Diest, P.J., Polónia, A., Curado, M., Temprana-Salvador, J., Zlobec, I., Purqueras, E., Weis, C.A., Matias-Guiu, X., Schirmacher, P., Ryška, A.: Digital transformation of pathology – the European Society of Pathology expert opinion paper. Virchows Archiv **487**(5), 971–981 (2025). https://doi.org/10.1007/s00428-025-04090-w
4. Fatima, G., Alhmadi, H., Mahdi, A.A., Hadi, N., Fedacko, J., Magomedova, A., Parvez, S., Raza, A.M.: Transforming Diagnostics: A Comprehensive Review of Advances in Digital Pathology. Cureus **16**(10), e71890 (2024). https://doi.org/10.7759/cureus.71890
5. Haghighat, M., Browning, L., Sirinukunwattana, K., Malacrino, S., Khalid Alham, N., Colling, R., Cui, Y., Rakha, E., Hamdy, F.C., Verrill, C., Rittscher, J.: Automated quality assessment of large digitised histology cohorts by artificial intelligence. Scientific Reports **12**(1), 5002 (2022). https://doi.org/10.1038/s41598-022-08351-5
6. Janowczyk, A., Zuo, R., Gilmore, H., Feldman, M., Madabhushi, A.: HistoQC: An Open-Source Quality Control Tool for Digital Pathology Slides. JCO Clinical Cancer Informatics **3**, 1–7 (2019). https://doi.org/10.1200/CCI.18.00157

7. Kahaki, S., Webber, A.R., Zamzmi, G., Subbaswamy, A., Deshpande, R., Badano, A.: HistoART: Histopathology Artifact Detection and Reporting Tool (2025). https://doi.org/10.48550/arXiv.2507.00044
8. Kanwal, N., Khoraminia, F., Kiraz, U., Mosquera-Zamudio, A., Monteagudo, C., Janssen, E.A.M., Zuiverloon, T.C.M., Rong, C., Engan, K.: Equipping computational pathology systems with artifact processing pipelines: a showcase for computation and performance trade-offs. BMC Medical Informatics and Decision Making **24**(1), 288 (2024). https://doi.org/10.1186/s12911-024-02676-z
9. Kanwal, N., López-Pérez, M., Kiraz, U., Zuiverloon, T.C.M., Molina, R., Engan, K.: Are you sure it's an artifact? Artifact detection and uncertainty quantification in histological images. Computerized Medical Imaging and Graphics **112**, 102321 (2024). https://doi.org/10.1016/j.compmedimag.2023.102321
10. Kothari, S., Phan, J.H., Wang, M.D.: Eliminating tissue-fold artifacts in histopathological whole-slide images for improved image-based prediction of cancer grade. Journal of Pathology Informatics **4**(1), 22 (2013). https://doi.org/10.4103/2153-3539.117448
11. Kumar, N., Gupta, R., Gupta, S.: Whole Slide Imaging (WSI) in Pathology: Current Perspectives and Future Directions. Journal of Digital Imaging **33**(4), 1034–1040 (2020). https://doi.org/10.1007/s10278-020-00351-z
12. Litjens, G., ASAP contributors: ASAP: Automated Slide Analysis Platform. https://github.com/computationalpathologygroup/ASAP (2022)
13. Patil, A., Diwakar, H., Sawant, J., Kurian, N.C., Yadav, S., Rane, S., Bameta, T., Sethi, A.: Efficient quality control of whole slide pathology images with human-in-the-loop training. Journal of Pathology Informatics **14**, 100306 (2023). https://doi.org/10.1016/j.jpi.2023.100306
14. Patil, A., Jain, G., Diwakar, H., Sawant, J., Bameta, T., Rane, S., Sethi, A.: Semantic segmentation based quality control of histopathology whole slide images. arXiv preprint arXiv:2410.03289 (oct 2024), https://arxiv.org/abs/2410.03289
15. Schömig-Markiefka, B., Pryalukhin, A., Hulla, W., Bychkov, A., Fukuoka, J., Madabhushi, A., Achter, V., Nieroda, L., Büttner, R., Quaas, A., Tolkach, Y.: Quality control stress test for deep learning-based diagnostic model in digital pathology. Modern Pathology **34**(12), 2098–2108 (2021). https://doi.org/10.1038/s41379-021-00859-x
16. Scotto, M., Patti, R., L'imperio, V., Fraggetta, F., Molinari, F., Salvi, M.: SlideInspect: From Pixel-Level Artifact Detection to Actionable Quality Metrics in Digital Pathology. International Journal of Imaging Systems and Technology **36**(1), e70292 (2026). https://doi.org/10.1002/ima.70292
17. Shakarami, A., Nicolè, L., Terreran, M., Dei Tos, A.P., Ghidoni, S.: TCNN: A Transformer Convolutional Neural Network for artifact classification in whole slide images. Biomedical Signal Processing and Control **84**, 104812 (2023). https://doi.org/10.1016/j.bspc.2023.104812
18. Smit, G., Ciompi, F., Cigéhn, M., Bodén, A., van der Laak, J., Mercan, C.: Quality control of whole-slide images through multi-class semantic segmentation of artifacts. In: Medical Imaging with Deep Learning (2021), https://openreview.net/forum?id=7EZ4JOtlRl
19. Szymanski, L., Lewicki, S., Markiewicz, T., Cierniak, S., Tassan, J.P., Kubiak, J.Z.: siRNA-mediated MELK knockdown induces accelerated wound healing with increased collagen deposition. International Journal of Molecular Sciences **24**(2), 1326 (2023). https://doi.org/10.3390/ijms24021326, published: 10 Jan. 2023

20. Taqi, S.A., Sami, S.A., Sami, L.B., Zaki, S.A.: A review of artifacts in histopathology. Journal of Oral and Maxillofacial Pathology **22**(2), 279 (2018). https://doi.org/10.4103/jomfp.JOMFP_125_15
21. Wang, N.C., Kaplan, J., Lee, J., Hodgin, J., Udager, A., Rao, A.: Stress testing pathology models with generated artifacts. Journal of Pathology Informatics **12**, 54 (12 2021). https://doi.org/10.4103/jpi.jpi_6_21
22. Wei, J.W., Tafe, L.J., Linnik, Y.A., Vaickus, L.J., Tomita, N., Hassanpour, S.: Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks. Scientific Reports **9**(1), 3358 (2019). https://doi.org/10.1038/s41598-019-40041-7
23. Weng, Z., Seper, A., Pryalukhin, A., Mairinger, F., Wickenhauser, C., Bauer, M., Glamann, L., Bläker, H., Lingscheidt, T., Hulla, W., Jonigk, D., Schallenberg, S., Bychkov, A., Fukuoka, J., Braun, M., Schömig-Markiefka, B., Klein, S., Thiel, A., Bozek, K., Netto, G.J., Quaas, A., Büttner, R., Tolkach, Y.: GrandQC: A comprehensive solution to quality control problem in digital pathology. Nature Communications **15**(1), 10685 (12 2024). https://doi.org/10.1038/s41467-024-54769-y
24. Wu, H., Phan, J.H., Bhatia, A.K., Cundiff, C.A., Shehata, B.M., Wang, M.D.: Detection of blur artifacts in histopathological whole-slide images of endomyocardial biopsies. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 727–730 (2015). https://doi.org/10.1109/EMBC.2015.7318465
25. Xiang, Y., Liu, B., Rantalainen, M.: A mixture of experts (MoE) model to improve AI-based computational pathology prediction performance under variable levels of image blur. BMC Medical Imaging **25**(1), 407 (10 2025). https://doi.org/10.1186/s12880-025-01974-w