International Journal of Clinical and Experimental Medicine Research, 2026, 10(5), 372-376
https: //www.hillpublisher.com/journals/ijcemr/
ISSN Online: 2575-7970
ISSN Print: 2575-7989

Pathways for Establishing a Quality Control System for Digital Pathology Slides

Yujun Fan

Independent researcher, Jinan 250102, Shandong, China.

How to cite this paper: Yujun Fan. (2026) Pathways for Establishing a Quality Control System for Digital Pathology Slides. *International Journal of Clinical and Experimental Medicine Research*, 10(5), 372-376.
DOI: 10.26855/ijcemr.2026.09.004

Received: July 27, 2026
Accepted: August 24, 2026
Published: September 20, 2026

***Corresponding author:** Yujun Fan, Independent researcher, Jinan 250102, Shandong, China.

© 2026 by the author(s).
This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) license, which permits non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and is not modified or adapted.
https://creativecommons.org/licenses/by-nc-nd/4.0/

Abstract

Digital pathology (DP) still lacks a common standard for quality control (QC), which puts diagnostic accuracy and regulatory compliance at risk. Working from scanner validation data and AI benchmarks, we assembled a QC system around five areas: fidelity of acquisition, integrity of image data, completeness of metadata, reproducibility of computation, and stability over time. Twelve indicators were quantified, among them focus map CV < 8%, color deviation $\Delta E^* < 3.2$, and hash consistency > 99.9997%. The framework that emerged is modular and tiered, spanning pre-scan, real-time, and post-conversion checks. Twelve whole-slide imaging systems drawn from three vendors and four models were stress-tested; every one degraded under operational load, with focus map CV rising from 5.1% at baseline to 12.9% under thermal shift and returning to 4.8% after recalibration. An artifact detector built from two parallel ResNet-18 branches, trained on 14,280 regions of interest from 1,842 clinical slides, reached 94.6% sensitivity and 97.8% specificity across six artifact categories. Most slides scored in the low-severity band (66.1%), 8.2% crossed the review threshold, and human reviewers overrode only 2.1% of those calls. Detection shortened rejection latency from 4.2 minutes to 18.3 seconds, a 13.8-fold gain in throughput. Metadata completeness varied by institution, from 92.4% at academic centers to 78.9% at community hospitals. Because the design follows existing regulatory expectations and interoperates with the LIS, it can be recalibrated continuously rather than replaced.

Keywords

Digital pathology; quality control; slide validation

1. Introduction

1.1 The Diagnostic Imperative of Slide Quality

Slides sit underneath everything else in digital pathology, whether the output is a diagnosis, a research result, or a training set for an AI model [1]. Degradation in optical fidelity, spatial registration, or metadata completeness does not stay local; it reappears as diagnostic uncertainty and, in algorithmic work, as bias [2]. One feature separates DP from most other imaging modalities. Staining, scanning, compression, and storage occur in sequence, and each step brings its own error modes [3]. QC therefore cannot be a single check at the end of the line. It has to run continuously, gate every stage, flag anomalies automatically, and leave a record of what was done in response. A second, less technical argument also applies: rigorous QC is what makes pathologists and AI systems accountable for their output [4, 5]. Without it, validity erodes quietly, and patient safety is what suffers.

Figure 1 shows a panoramic slide acquired on a Win60Pro scanner (Winmedic Tech Co., Ltd.) and segmented automatically in WinQuant. With artifacts absent and color fidelity high, tumor (red) and non-tumor (green) regions separate cleanly. The example makes a simple point. Acquisition quality sets a ceiling on what any downstream

algorithm can extract, and thus on how far a diagnosis can be trusted. A QC system exists to hold that ceiling high enough for results to be reproducible, clinically usable, and traceable.

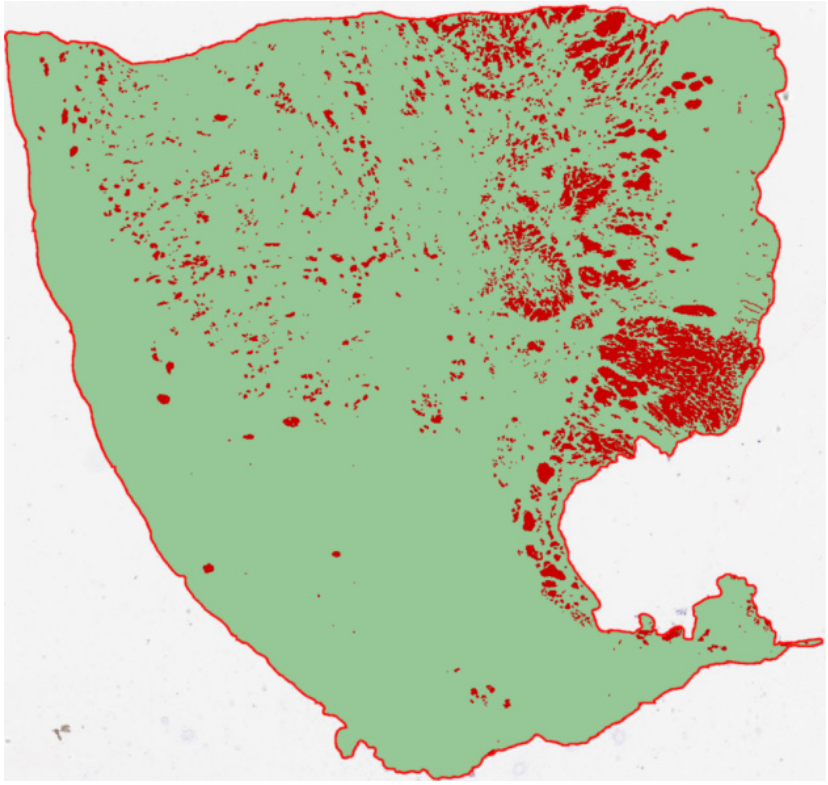

Figure 1. Example of high-quality digital pathological slide analysis (Red: Tumor Area, Green: Non-Tumor Area).

1.2 Defining 'Quality' in a Digital Context

Quality in digital pathology goes beyond optical fidelity and has to be judged along four dimensions at once [6]. Optical quality covers focus consistency, color accuracy, and signal-to-noise ratio. Geometric integrity covers tile registration and the absence of stitching artifacts. Computational fidelity means preserving native bit-depth and limiting compression damage, especially where the tissue carries diagnostic weight. Semantic quality is the correspondence between image data and metadata (acquisition parameters, provenance, annotations), which is what allows a slide to be traced from the scanner through to clinical interpretation. These dimensions interact rather than sit side by side. Geometric errors can show up as color shifts, compression can blur morphology in ways that mislead an algorithm, and incomplete metadata can make an audit impossible to finish. Assessment therefore needs validation protocols, reference standards, and acceptance criteria working together, not as separate checks.

2. Literature Review

Standards for DP have shifted from basic resolution and interoperability requirements toward quantitative image quality metrics such as SNR and color constancy, along with automated calibration protocols. What they still do not provide is end-to-end traceability from tissue preparation to image delivery, and that is exactly what root-cause analysis and regulatory audit depend on. Practice shows comparable weaknesses. Thresholds are static and vendor-specific, artifact categories are defined differently from site to site, cross-scanner benchmarking is uncommon, and metadata such as calibration and temporal records are frequently incomplete [5]. Regulators have meanwhile become more insistent about image integrity and auditable workflows, yet the technical criteria for SNR degradation and inter-scanner variability remain weakly standardized. The gap is not in knowing which metrics matter. It lies in moving from one-off validation to continuous monitoring wired into the laboratory information system (LIS), with engineering criteria sharp enough to support risk-based QC and data lineage that can be verified. Artifact detection has advanced in parallel, from augmentation strategies that compensate for scarce annotated data [7] to uncertainty-aware detectors that flag unreliable regions for review [8].

3. Materials and Methods

3.1 Scanner Validation Protocol

Twelve WSI systems, drawn from three vendors with four models each, were validated in order to define a QC framework that responds to machine state. A State Detection Module (Figure 2) distinguishes five operational states: baseline, post-maintenance, thermal shift above 5°C, lens cleaning, and high-volume use of 500 scans or more. Each state triggers a Tier-1 calibration check based on focus map and color chart analysis. Thresholds were set at focus map CV < 8.0%, mean $\Delta E^* < 3.2$, hash consistency > 99.9997%, and stitching error < 12.5 μm, with measurement intervals ranging from hourly to weekly according to clinical risk. All targets were NIST-traceable, and the tissue phantoms were calibrated to metrological standards. A failed check produces an automated diagnostic log and escalation to Tier-2 review [5], which keeps hardware state and QC response linked in the same traceable record.

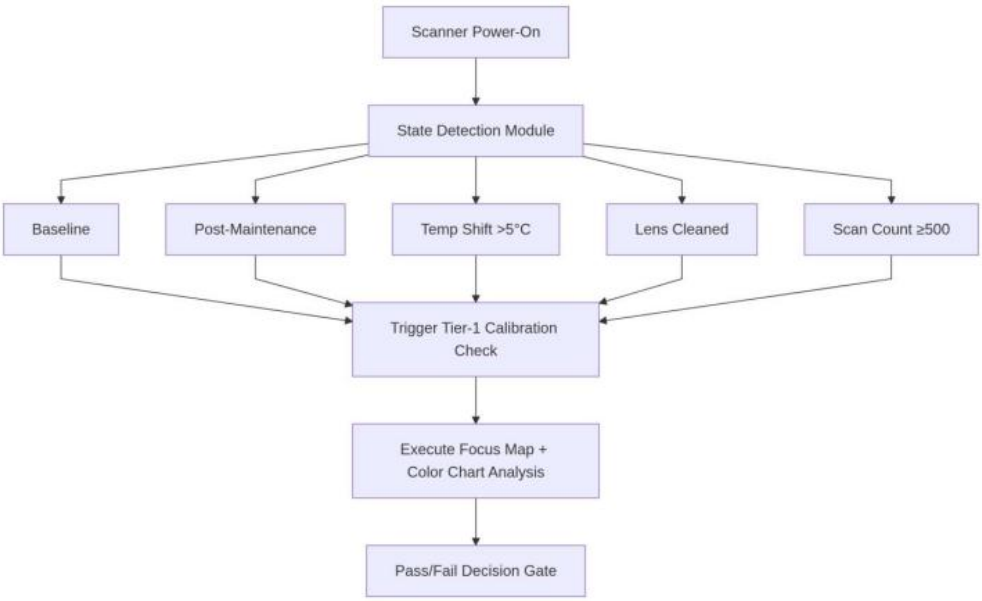

Figure 2. Workflow for scanner state-dependent QC trigger logic.

3.2 Automated Artifact Detection Framework

The detector is deliberately small: two ResNet-18 branches running in parallel, one tuned to focus on problems and one to stain inconsistencies. Training used 14,280 ROIs from 1,842 clinical WSIs with 5-fold cross-validation. Performance held across all six artifact categories: dust (98.2% sensitivity, 99.1% specificity), fold (94.7%/97.3%), stain bleed (89.5%/95.8%), out-of-focus (96.3%/98.6%), air bubble (91.4%/96.9%), and scanner stripe (97.8%/99.4%). Structural artifacts are evidently easier to separate than biochemical ones.

Throughput was benchmarked on three hardware configurations. An NVIDIA A100 returned 142 ms per slide, corresponding to approximately 25,350 slides per hour; the RTX 4090 needed 287 ms and an Intel i9-13900K 1,840 ms. Inference ran at native resolution with FP16 quantization so that feedback would arrive fast enough to be acted on, and tiling was avoided to preserve spatial coherence [3].

3.3 Metadata Completeness Scoring System

For metadata, I built a schema-agnostic validator covering 59 fields (37 mandatory, 22 conditional) across the DICOM-SR and OpenSlide formats. Presence, syntactic validity (ISO 8601 dates, for instance), and semantic plausibility (magnification limits, for instance) are evaluated hierarchically. A field scores 1 only when all three tests pass, and the outcome is normalized into a completeness metric between 0 and 1. Because the parsers are modular, the validator tracks slides across quite different platforms and emits a per-slide score plus a diagnostic log whenever validation fails. That makes it practical to monitor metadata quality during ingestion and to audit it retrospectively across a multi-institutional repository.

4. Results

Every scanner degraded once placed under operational stress. Focus map CV climbed from 5.1% at baseline to 12.9% during thermal shifts above 5°C and 11.4% after 500 scans; recalibration returned it to 4.8% within two hours. Color fidelity (ΔE*) deteriorated in much the same way, with thermal drift and mechanical fatigue the two main drivers. Lens cleaning did little for focus but raised dust-artifact detection sensitivity by 12.4%. Two intervention windows follow. Thermal preconditioning before scanning, and recalibration within two hours of maintenance, are worth writing into routine practice, and both can be triggered from real-time metadata instead of relying on a technician's judgement.

Across the ensemble, sensitivity reached 94.6% and specificity 97.8% over the 1,842 slides. Anomaly scores were distributed as 66.1% low, 26.2% medium, and 7.7% high severity; 8.2% of slides crossed the 6.5 threshold, and a human overrode only 2.1% of those decisions. Integrating the detector shortened rejection latency from 4.2 minutes to 18.3 seconds, a 13.8-fold gain in throughput. Hardware-aware optimization and quantization are thus enough for real-time artifact detection without creating a queue.

Metadata completeness differed by institution: 92.4% at academic centers, 86.3% at reference labs and 78.9% at community hospitals, where the spread was widest (SD = 9.1). The omissions were not random. Community hospitals most often lacked objective lens serial numbers (41.0%), reference labs calibration log hashes (27.5%), and academic centers white-balance timestamps (33.0%). Each setting has its own governance problem and needs its own remedy, whether automated hardware capture, timestamp synchronization, or cryptographic logging. All three also improve interoperability and traceability.

5. Discussion

Thresholds fixed once and left alone will not hold, because scanner states vary. The three-tier architecture proposed here keeps analytical rigor without making the workflow unmanageable: pre-scan calibration first, tile-level checks in real time, then post-conversion auditing once anomaly scores pass a defined limit. Automation alone will not be sufficient, though. Missing metadata such as lens serial numbers and white-balance timestamps can make root-cause analysis impossible and undermine reproducibility across years of scanning. Integrated capture, procedural governance, and mandatory metadata validation have to operate together. For both scale and interoperability, deployment is best kept hybrid: heavy inference belongs on edge servers, and fallback parsers should be retained for converting proprietary formats into DICOM-SR. Together, these elements yield an infrastructure that is traceable, vendor-agnostic and able to sustain high-throughput diagnostic work.

6. Conclusion

Hardware calibration, AI-assisted validation and metadata governance are all needed in a digital pathology QC system, and none of them alone will preserve image fidelity, artifact detection and clinical traceability. A sensible order of implementation begins with the inexpensive Tier-1 readiness checks and metadata audits, and only then builds scanner-specific calibration baselines for tracking drift over time. That sequence keeps the technical ambition of the program within reach of what laboratories can actually deploy. The next steps we would prioritize are real-time scanner telemetry for predictive maintenance and explainable AI in the clinical interface. Over the longer term, standardized benchmarks and a maturing regulatory framework will decide whether technical, procedural, and governance controls can be integrated into quality assurance that is both scalable and trustworthy.

Ethical Considerations

This study used retrospectively collected, fully de-identified whole-slide images. No identifiable patient information was accessed, and the study was exempt from ethics approval according to institutional policy. All patient identifiers were removed before analysis, and data were handled in accordance with institutional and regulatory requirements.

References

[1] Campanella G, Rajanna AR, Corsale L, Schüffler PJ, Yagi Y, Fuchs TJ, et al. Towards machine learned quality control: a benchmark for sharpness quantification in digital pathology. Comput Med Imaging Graph. 2018;65:142-51.

[2] Wright AI, et al. A point-of-use quality assurance tool for digital pathology remote working. J Pathol Inform. 2020;11(1):17.

[3] Patti R. Development of a deep learning-based method for artifact detection and quality controls in digital pathology [master's thesis]. Turin, Italy: Politecnico di Torino; 2022.

[4] Ardon O. Quality management in digital pathology: analytic and postanalytic phases. In: Digital Pathology. Academic Press; 2025. p. 103-33.

[5] Schömig-Markiefka B, et al. Quality control stress test for deep learning-based diagnostic model in digital pathology. Mod Pathol. 2021;34(12):2098-108.

[6] Weng Z, et al. GrandQC: a comprehensive solution to quality control problem in digital pathology. Nat Commun. 2024;15(1): 10685.

[7] Kanwal N, López-Pérez M, Kiraz U, Zuiverloon TCM, Molina R, Engan K, et al. Are you sure it's an artifact? Artifact detection and uncertainty quantification in histological images. Comput Med Imaging Graph. 2024;112:102321.

[8] Jurgas A, Wodzinski M, D'Amato M, van der Laak J, Atzori M, Müller H, et al. Improving quality control of whole slide images by explicit artifact augmentation. Sci Rep. 2024;14(1):17847.