GMR
The Original

COMPARATIVE DIAGNOSTIC CONCORDANCE AND TURNAROUND TIME OF ARTIFICIAL INTELLIGENCE-ASSISTED AND CONVENTIONAL HISTOPATHOLOGY: A PROSPECTIVE OBSERVATIONAL STUDY

Dr. Karthick. G[1], Dr. Govindarajan. R[2]*, Dr. Karthik. J[3]

[1]Assistant Professor Department Of Pathology, Shri Sathya Sai Medical College and Research Institute (Sri Balaji Vidyapeeth University (DU), Chennai campus), Ammapettai, Nellikuppam Post, Chengalpattu District, Tamilnadu, 603108, India karthikhbk27@gmail.com
[2]Professor & Head, Department of Pathology, Shri Sathya Sai Medical College & Research Institute, (Sri Balaji Vidyapeeth University (DU), Chennai campus), Ammapettai, Nellikuppam Post, Chengalpattu District Tamilnadu, 603108, India gjhospital@yahoo.com
[3]Assistant professor, Department of Pathology, Sri Venkateshwaraa Medical College Hospital and Research Centre, Pondicherry university, 13-A, Pondy - Villupuram Main Road, Ariyur, Puducherry, 605 102, India.
drkarthikjms@gmail.com
Corresponding Author*: Dr. Govindarajan. R, E-mail: gjhospital@yahoo.com

ABSTRACT

Background: Whole-slide imaging has created a practical route for artificial intelligence (AI) to support histopathological interpretation. Its value in routine diagnostic practice depends not only on algorithmic accuracy but also on whether it improves pathologist concordance and reduces reporting time across clinically varied specimens.

Objectives: To compare case-level diagnostic concordance with a final integrated reference diagnosis and the time required for AI-only output, conventional pathologist diagnosis, and pathologist diagnosis augmented by AI.

Materials and Methods: This prospective observational comparative study included 50 consecutive histopathology cases evaluated in the Department of Pathology, Shri Sathya Sai Medical College and Research Institute, Tiruporur, during 2025–2026. Routine slides were digitally scanned. An AI-assisted diagnostic system generated a top-ranked diagnosis and confidence score. Two pathologists initially interpreted the cases without AI and, after a washout period, re-evaluated the digital slides with AI support to record a consensus diagnosis. The final integrated histopathological diagnosis, incorporating clinical correlation and ancillary investigations where required, served as the reference standard. Exact concordance was compared using McNemar's test; paired diagnostic times were analysed using the Friedman and Wilcoxon signed-rank tests.

Results: The mean age was 52.08±13.19 years, and 27 (54.0%) patients were female. The reference diagnoses were malignant in 25 (50.0%), benign in 15 (30.0%), inflammatory in 6 (12.0%), and premalignant in 4 (8.0%) cases. Exact concordance with the reference standard was 100% for the AI-only output and AI-augmented pathologist diagnosis, compared with 68.0% for conventional pathologist diagnosis (34/50; exact McNemar $p<0.001$). All 16 conventional discordances were concordant after AI augmentation. Mean diagnostic time was 0.51±0.10 minutes for AI output, 12.58±2.84 minutes for conventional interpretation, and 6.50±1.50 minutes for AI-augmented interpretation. AI assistance reduced mean pathologist review time by 48.3% ($p<0.001$). As no discordance occurred in either AI-related arm, AI error patterns and calibration of confidence scores against diagnostic error could not be evaluated.

Conclusion: Within this small, heterogeneous single-centre cohort, AI-supported interpretation was associated with higher exact diagnostic concordance and substantially shorter review time than conventional interpretation alone. The perfect concordance in both AI-related arms should be regarded as hypothesis-generating because selection effects, reference-standard dependence, and the absence of external validation may have influenced the observed estimates.

KEYWORDS: computational pathology; decision support; digital slides; diagnostic agreement; reporting efficiency; whole-slide imaging.

INTRODUCTION

Histopathological diagnosis remains a pattern-recognition task performed within a demanding clinical workflow. The introduction of whole-slide imaging has shifted part of that workflow from optical microscopy to high-resolution digital review. Large multicentre studies have shown that digital whole-slide interpretation can achieve diagnostic performance non-inferior to conventional light microscopy across biopsies, resections, special stains, and diverse organ systems.[1,2] This digital foundation is important because most contemporary AI systems in pathology depend on scanned slides rather than direct microscopic viewing.

Deep-learning models can identify morphological patterns, prioritise suspicious regions, classify lesions, and quantify features that are otherwise time-consuming or vulnerable to observer variation. Clinical-grade algorithms have demonstrated strong performance in tumour detection and classification across prostate, breast, gastrointestinal, and lung pathology.[3–6] Their practical role, however, is more appropriately framed as decision support than autonomous replacement. The pathologist must still integrate morphology with specimen adequacy, clinical information, immunohistochemistry, molecular findings, and the consequences of diagnostic uncertainty.

Reader studies have begun to show a measurable human–AI interaction. In lymph-node assessment and prostate biopsy interpretation, algorithm assistance has improved sensitivity or agreement while reducing reading time.[7–10] Yet most published evaluations focus on a single disease, one organ system, or a narrowly defined diagnostic task. Routine pathology laboratories face a more heterogeneous case mix, including benign, inflammatory, premalignant, and malignant lesions from multiple anatomical sites.

Evidence from Indian institutions also indicates that computational methods can be developed for locally relevant pathology tasks, including urothelial carcinoma grading and automated biomarker assessment.[11,12] Even so, implementation requires local validation, attention to staining and scanning variability, and a clear distinction between algorithm output and the authorised final diagnosis. The present study compared AI-only output, conventional pathologist interpretation, and pathologist interpretation augmented by AI against an integrated reference diagnosis, while also examining the time required by each diagnostic approach.

MATERIALS AND METHODS

Study design and setting

A prospective observational comparative study was conducted in the Department of Pathology, Shri Sathya Sai Medical College and Research Institute, Tiruporur, Tamil Nadu.

Study period

The study was carried out during 2025–2026.

Study population and case selection

Fifty consecutive histopathology cases received during the study period were included. The case series comprised biopsies, resections, cytology cell blocks, and other routinely processed tissue specimens. Cases represented benign, inflammatory, premalignant, and malignant diagnoses from multiple anatomical sites.

Slide preparation and digitisation

Routine diagnostic slides were prepared according to the departmental workflow and digitally scanned to generate whole-slide images. Ancillary investigations, including special stains, immunohistochemistry, microbiological findings, or other relevant tests, were incorporated when clinically indicated. Scanner model, optical resolution, file format, and image-quality rejection criteria were not available for inclusion in the study record.

Diagnostic assessment

Three diagnostic assessments were recorded for each case. First, the AI-assisted diagnostic system generated a top-ranked diagnosis and a confidence score expressed as a percentage. The study log identified the system only by the local label ‘PathAI_v2.1’. A release date, manufacturer-issued version documentation, validated tissue spectrum, intended-use population, deployment configuration, training data, and indication-specific regulatory authorisation were not available in the study record. Accordingly, no claim is made that this label represents a commercially marketed or independently validated diagnostic product; it is reported solely as the AI decision-support tool evaluated in this cohort. Second, two pathologists independently evaluated the cases without AI support, and the case-level unassisted diagnosis recorded in the study log was used as the conventional diagnostic arm. Third, following a washout period, the same pathologists reviewed the digital slides with access to the AI output and recorded an AI-augmented consensus diagnosis. Reader-specific diagnoses were not separately available for interobserver analysis.

Reference standard

The final integrated diagnosis served as the reference standard. It was based on histomorphology together with clinical correlation and ancillary investigations where required. Concordance was defined as exact agreement of the study-arm diagnosis with the recorded reference diagnosis.

Outcome measures

The primary outcome was case-level exact diagnostic concordance for each of the three study arms. Secondary outcomes were time to diagnosis, the number and nature of conventional diagnostic discordances, correction of discordant diagnoses after AI assistance, and the distribution of AI confidence scores.

Statistical analysis

Continuous variables were summarised using mean±standard deviation and median with interquartile range (IQR). Categorical variables were expressed as frequency and percentage. Exact diagnostic concordance was reported with Wilson 95% confidence intervals. Paired differences in concordance were evaluated using the exact McNemar test. Diagnostic times across the three arms were compared using the Friedman test, followed

by paired Wilcoxon signed-rank tests. A two-sided p-value <0.05 was considered statistically significant. The analysis was exploratory and no adjustment for multiple pairwise comparisons was applied.

Ethical considerations

The study evaluated routinely processed diagnostic material within an observational workflow. AI-generated outputs were assessed as decision-support information and were not used as the sole basis for patient management. Patient identifiers were excluded from the analytical record, and confidentiality was maintained in accordance with institutional ethical principles.

RESULTS

Patient and diagnostic profile

The study included 50 cases from patients aged 28–76 years, with a mean age of 52.08±13.19 years. There were 27 (54.0%) females and 23 (46.0%) males. The reference standard classified 25 (50.0%) cases as malignant, 15 (30.0%) as benign, 6 (12.0%) as inflammatory, and 4 (8.0%) as premalignant (Table 1 and Figure 1). The mean AI confidence score was 92.64±4.24%, with a median of 93.0% (IQR 90.0–96.0%) and a range of 83–99%.

Table 1. Baseline demographic and diagnostic characteristics (n=50).

Characteristic	Value
Age, years	52.08±13.19 (28–76)
Sex, female	27 (54.0)
Sex, male	23 (46.0)
Malignant diagnosis	25 (50.0)
Benign diagnosis	15 (30.0)
Inflammatory diagnosis	6 (12.0)
Premalignant diagnosis	4 (8.0)
AI confidence score, %	92.64±4.24
AI confidence score, median (IQR)	93.0 (90.0–96.0)

Values are mean±standard deviation (range), median (interquartile range), or n (%), as applicable. AI: artificial intelligence.

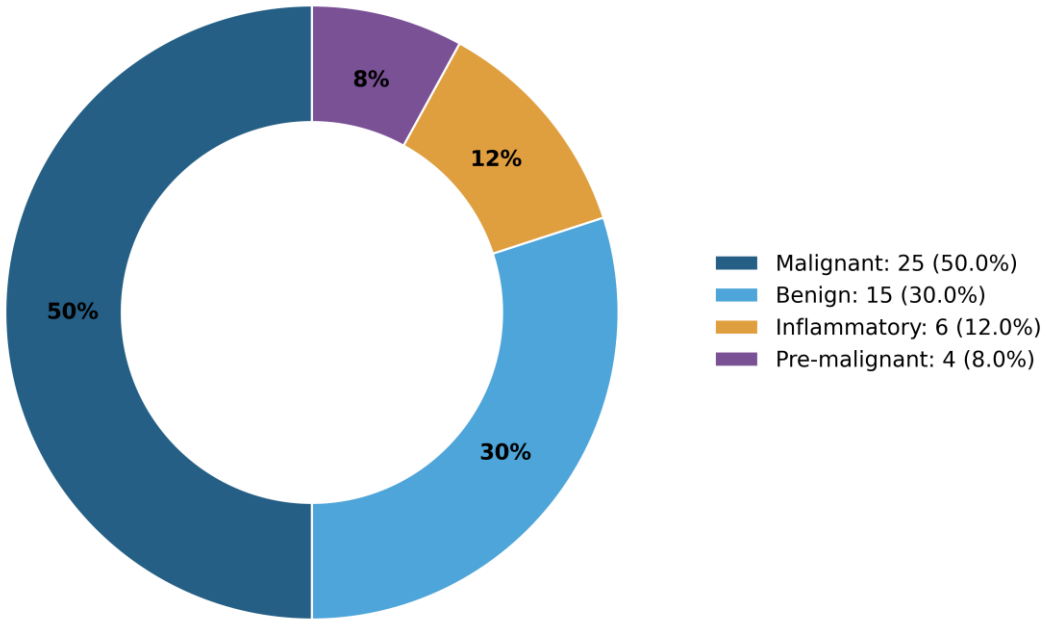

Figure 1. Distribution of final integrated diagnostic categories.

The donut chart displays the proportion and number of cases in each reference-diagnosis category.

Diagnostic concordance with the reference standard

The AI-only diagnosis agreed exactly with the reference standard in all 50 cases, corresponding to 100% concordance (95% CI 92.9–100). The conventional pathologist diagnosis was concordant in 34/50 cases, giving an accuracy of 68.0% (95% CI 54.2–79.2). Following AI augmentation, all 50 case-level consensus diagnoses

matched the reference standard (100%; 95% CI 92.9–100). Both the AI-only and AI-augmented arms showed significantly higher concordance than conventional interpretation alone (exact McNemar p<0.001). There was no observed difference between the AI-only and AI-augmented arms (Table 2 and Figure 2). Because both AI-related arms contained no discordant cases, false-positive and false-negative patterns, subgroup error rates, and confidence-score calibration could not be estimated.

Table 2. Exact diagnostic concordance with the final integrated reference diagnosis.

Diagnostic arm	Concordant cases, n/N	Accuracy (%)	95% CI	Comparison with conventional arm
AI-only output	50/50	100.0	92.9–100.0	p<0.001
Conventional pathologist diagnosis	34/50	68.0	54.2–79.2	Reference
Pathologist diagnosis augmented by AI	50/50	100.0	92.9–100.0	p<0.001

Accuracy was defined as exact case-level agreement with the recorded reference diagnosis. CI: confidence interval. Pairwise p-values were calculated using the exact McNemar test.

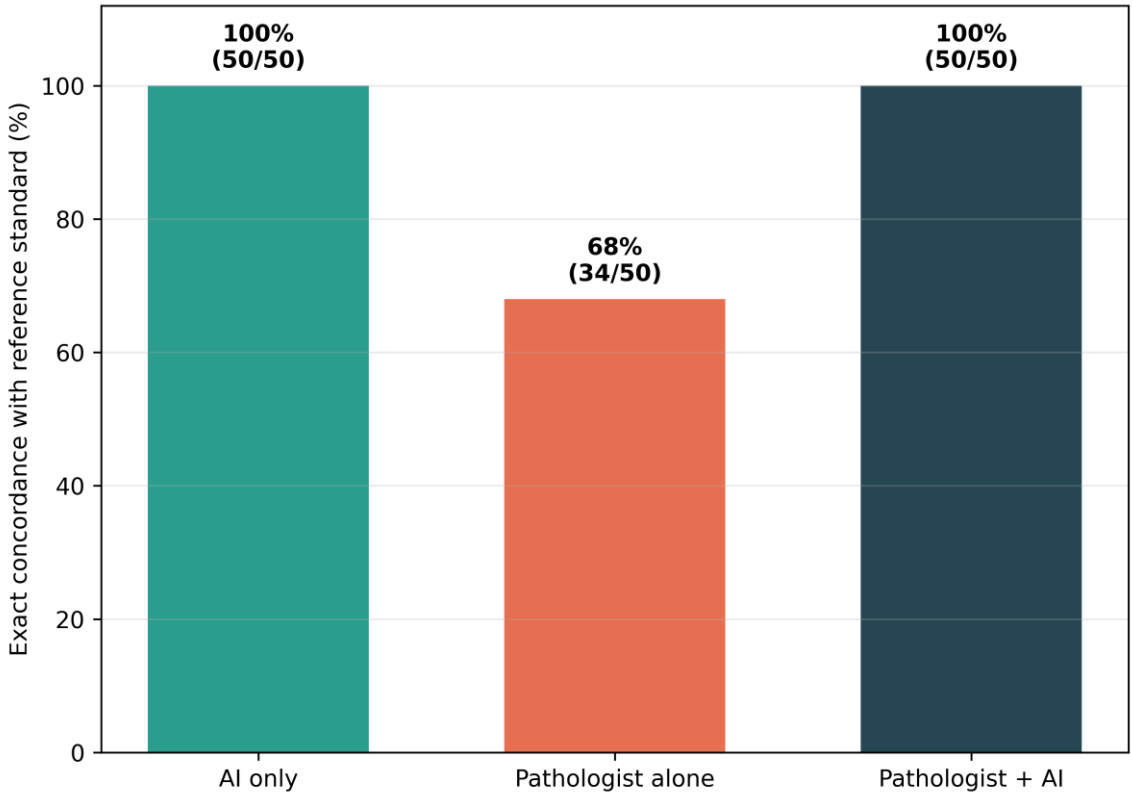

Figure 2. Exact diagnostic concordance across the three study arms.

Bars show the percentage and observed number of cases matching the final integrated reference diagnosis.

Nature of conventional diagnostic discordances

Conventional interpretation was discordant in 16 cases. Eleven of these discordances occurred in malignant cases, two in premalignant lesions, two in inflammatory conditions, and one in a benign lesion (Table 3). The discrepancies predominantly involved under-specification, grading or subtype refinement, or classification of diagnostically difficult lesions. Examples included atypical endometrial hyperplasia versus endometrioid adenocarcinoma, an atypical spindle-cell tumour versus leiomyosarcoma, adrenal cortical neoplasm versus adrenocortical carcinoma, atypical mesothelial proliferation versus malignant mesothelioma, and urothelial papilloma versus low-grade urothelial carcinoma. All 16 discordant conventional interpretations were concordant with the reference standard after AI-assisted review.

Table 3. Distribution of conventional diagnostic discordances by final diagnostic category.

Final category	Total cases	Conventional discordances	Discordance rate within category

Final category	Total cases	Conventional discordances	Discordance rate within category
Malignant	25	11	44.0%
Benign	15	1	6.7%
Inflammatory	6	2	33.3%
Pre-malignant	4	2	50.0%

Discordance denotes failure of the unassisted case-level diagnosis to match the recorded final integrated diagnosis exactly.

Table 4. Conventional discordant diagnoses corrected after AI-assisted review (n=16).

Case	Specimen	Conventional diagnosis without AI	Final integrated diagnosis
3	Endometrial biopsy	Atypical hyperplasia	Endometrioid adenocarcinoma
5	Colonic polyp	Tubular adenoma	Tubular adenoma with low-grade dysplasia
8	Gastric biopsy	Chronic gastritis	Chronic gastritis with H. pylori
10	Lung biopsy	Non-small cell carcinoma	Adenocarcinoma
13	Bone marrow biopsy	Plasma cell dyscrasia	Multiple myeloma
17	Duodenal biopsy	Celiac disease	Celiac disease (Marsh 3)
18	Soft tissue mass	Atypical spindle cell tumor	Leiomyosarcoma
26	Adrenal mass	Adrenal cortical neoplasm	Adrenocortical carcinoma
29	Pancreatic biopsy	Pancreatic adenocarcinoma	Pancreatic ductal adenocarcinoma
32	Prostate TURP	Prostatic adenocarcinoma	Prostatic adenocarcinoma Gleason 3+3
33	Thyroid lobectomy	Follicular neoplasm	Follicular adenoma
35	Lymph node	Metastatic carcinoma	Metastatic breast carcinoma
36	Pleural biopsy	Atypical mesothelial proliferation	Malignant mesothelioma
40	Bladder	Urothelial papilloma	Ta low-grade urothelial carcinoma
48	Ovary	Atypical serous proliferation	Borderline serous tumor
49	Kidney tumor	Renal cell carcinoma	Renal cell carcinoma clear cell type

In each listed case, the diagnosis recorded after AI-assisted review matched the final integrated reference diagnosis.

Time to diagnosis

Diagnostic time differed significantly across the three arms (Friedman χ^2=100.0, p<0.001). AI generated its top-ranked output in a mean of 0.51±0.10 minutes. Conventional interpretation required 12.58±2.84 minutes, while AI-augmented pathologist review required 6.50±1.50 minutes. The mean reduction from conventional to AI-augmented interpretation was 6.08 minutes, representing a 48.3% reduction in review time (Wilcoxon p<0.001). Pairwise differences between each of the three diagnostic arms were statistically significant (Table 5 and Figures 3–4).

Table 5. Comparison of time to diagnosis across the three study arms.

Diagnostic arm	Mean±SD, min	Median (IQR), min	Range, min	Pairwise comparison with conventional arm
AI-only output	0.51±0.10	0.50 (0.43–0.60)	0.3–0.7	p<0.001
Conventional pathologist diagnosis	12.58±2.84	12.00 (10.25–14.00)	8.0–20.0	Reference

Diagnostic arm	Mean±SD, min	Median (IQR), min	Range, min	Pairwise comparison with conventional arm
Pathologist diagnosis augmented by AI	6.50±1.50	6.00 (5.00–7.00)	4.0–10.0	p<0.001

Overall comparison: Friedman test p<0.001. Pairwise comparisons used the Wilcoxon signed-rank test. AI: artificial intelligence; IQR: interquartile range; SD: standard deviation.

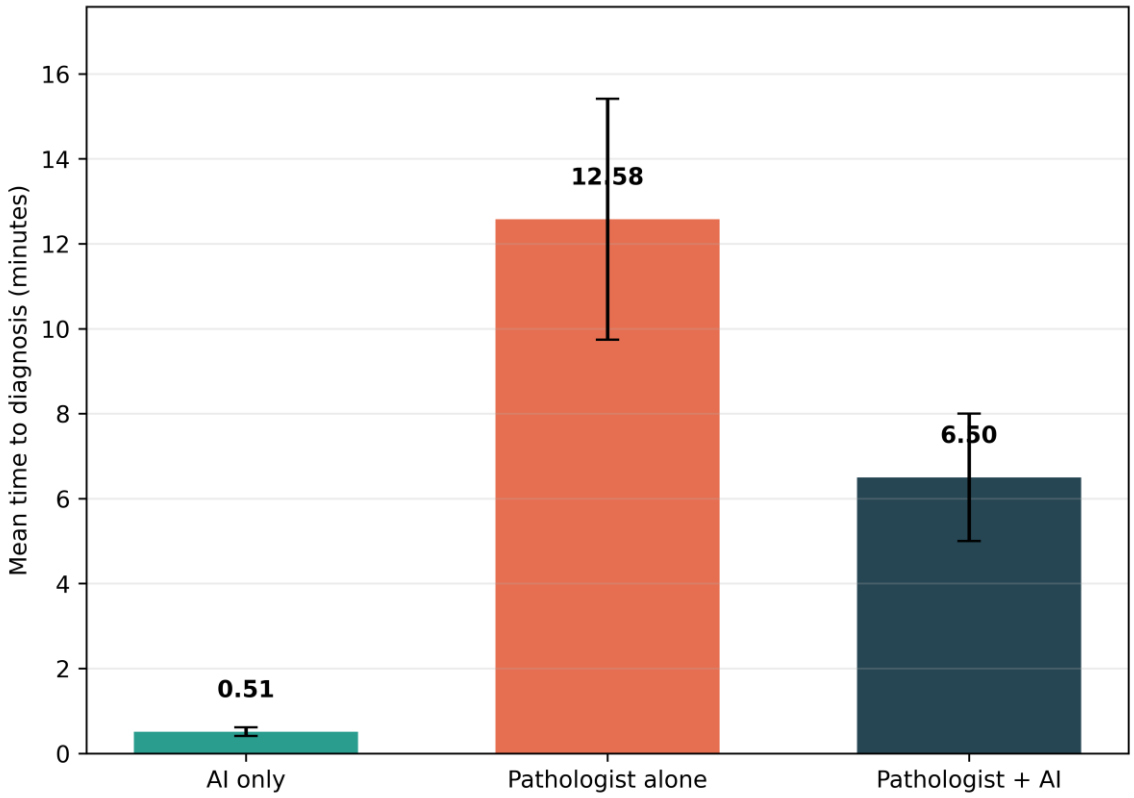

Figure 3. Mean time to diagnosis across the three study arms.
Bars display mean minutes; error bars represent standard deviation.

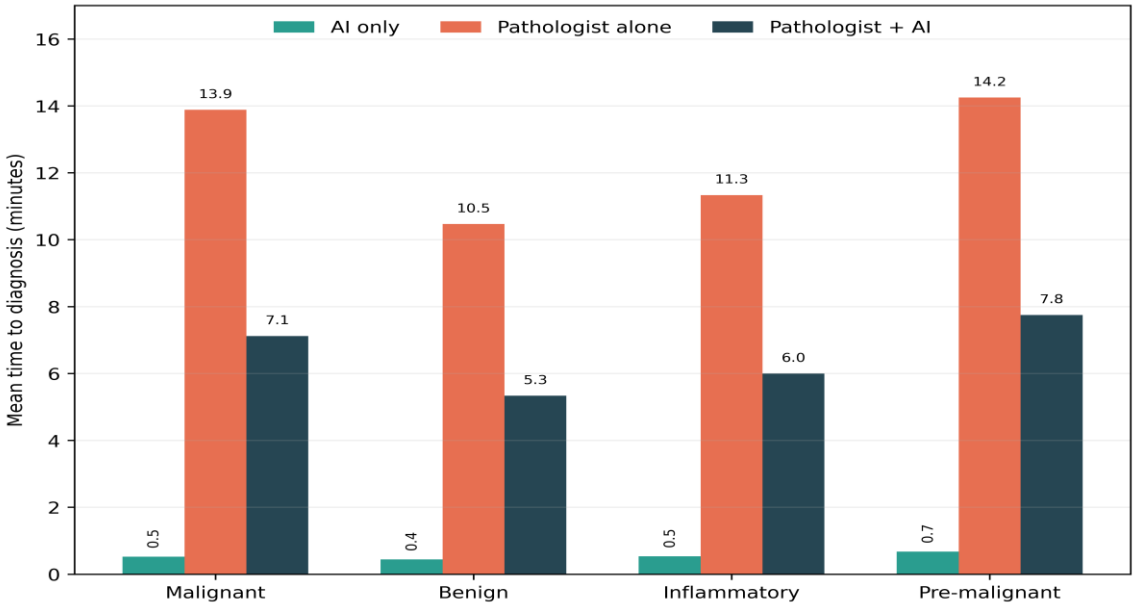

Figure 4. Mean diagnostic time according to final diagnostic category.
Grouped bars compare AI-only output, conventional pathologist review, and AI-augmented pathologist review within each category.

DISCUSSION

This prospective comparative study found a marked difference between conventional case-level interpretation and the two AI-related diagnostic arms. Exact agreement with the integrated reference diagnosis was 68% for the conventional arm, while the AI-only output and AI-augmented pathologist diagnosis agreed in all 50 cases. AI assistance also reduced mean pathologist review time by almost one-half. These findings support the operational premise that decision-support tools may help pathologists refine classification and reach a diagnosis more efficiently. At the same time, the perfect concordance observed in both AI arms is unusually high and should not be read as evidence that a general-purpose system can replace specialist interpretation across organ systems.

The digital component of the workflow is itself important. Large randomised non-inferiority studies have established that whole-slide images can support primary diagnosis across a broad surgical pathology case mix.[1,2] The present study therefore examined AI within a workflow that is already technically compatible with routine diagnostic interpretation. However, equivalence between digital and optical viewing does not automatically validate the AI layer. An algorithm must be assessed for its intended task, specimen types, scanners, staining processes, and local case prevalence.

The observed improvement with AI augmentation is directionally consistent with reader studies. Steiner et al. showed that deep-learning assistance improved pathologist detection of breast cancer micrometastases and reduced review time.[7] Retamero et al. later reported a 55% reduction in mean reading time, together with improved sensitivity for some readers during AI-assisted lymph-node assessment.[8] In prostate pathology, AI-supported interpretation has also increased cancer detection and improved grading agreement.[9,10] The 48.3% time reduction recorded here is similar in magnitude, although direct comparison is limited by differences in case mix, reader design, timing definitions, and algorithmic task.

Most of the unassisted discordances were not simple benign–malignant reversals. They involved refinement of grade, subtype, aetiology, or a diagnostically difficult borderline category. Examples included non-small-cell carcinoma versus adenocarcinoma, plasma-cell dyscrasia versus multiple myeloma, follicular neoplasm versus follicular adenoma, and metastatic carcinoma versus metastatic breast carcinoma. This pattern suggests that the AI output may have functioned as a structured second opinion that narrowed a differential diagnosis. Yet the study did not record reader-level reasoning, ancillary test requests, or whether the AI suggestion altered the order in which features were examined. It is therefore not possible to determine the precise mechanism of improvement.

A notable feature is the heterogeneous spectrum of tissues and diagnoses. Many landmark computational pathology studies have achieved high performance within tightly specified tasks, such as prostate cancer grading, lymph-node metastasis detection, or gastrointestinal tumour classification.[3–6] Pan-organ evaluation is more demanding because morphology, diagnostic thresholds, and relevant ancillary tests differ substantially. The present results demonstrate performance only within these 50 selected cases. They do not establish external validity for unrepresented diagnoses, rare entities, low-quality slides, cytology, frozen sections, or cases with conflicting ancillary data.

The mean AI confidence score exceeded 92%, but confidence values were not calibrated against error probability because no AI discordances occurred. High confidence should never be interpreted as equivalent to diagnostic certainty unless calibration has been validated on independent local data. Digital pathology algorithms can be sensitive to focus, tissue folds, staining variation, compression, scanner characteristics, and other preanalytical artefacts.[13] Task-specific studies have also demonstrated deep-learning classification of non-small-cell lung cancer and automated detection and quantification of invasive breast carcinoma on whole-slide images.[14,15] Such performance, however, cannot be assumed to transfer to a heterogeneous pan-organ cohort without task-specific and local validation. These considerations are especially relevant in laboratories transitioning to digital workflows, where variations in section quality and staining may be more consequential than differences in model architecture.

The findings also have practical relevance for Indian teaching hospitals. Digital pathology can facilitate remote consultation, teaching, workload distribution, and access to subspecialty support. Locally developed work has already shown the feasibility of AI-based classification of urothelial carcinoma and automated Ki-67 assessment in South Asian settings.[11,12] Still, implementation requires secure data governance, reliable image storage, institutional validation, defined responsibility for final sign-out, and contingency pathways when the algorithm and pathologist disagree. AI should remain an adjunct under pathologist oversight, particularly where the model's intended use and regulatory status are narrowly defined.

LIMITATIONS

The study was limited by its small sample size, single-centre setting, and heterogeneous but non-exhaustive case spectrum. Although cases were enrolled consecutively, the 50-case series may not represent the full range of diagnostic difficulty, uncommon entities, or poor-quality material encountered in routine practice. The conventional non-concordant group contained only 16 cases, and both AI-related arms showed perfect concordance, producing effect estimates that may be inflated by case selection, spectrum effects, or reference-standard dependence. Reader-specific diagnoses were not retained, preventing analysis of interobserver

agreement, individual reader improvement, and differential benefit according to reader experience. The duration of the washout period was not documented; consequently, recall, carry-over, and order effects cannot be excluded. Scanner model, image-resolution parameters, file format, slide-quality exclusions, and the exact ancillary-test pathway were unavailable, limiting technical reproducibility. The absence of verified software release documentation, intended-use population, and validated tissue scope further prevents exact replication or generalisation to other diagnostic settings. The reference diagnosis incorporated histopathology and ancillary information, but blinding of final adjudication to preceding AI-related outputs could not be confirmed, so incorporation or reference-standard bias remains possible. Timing captured diagnostic review rather than total laboratory turnaround time and did not include scanning, file transfer, quality control, ancillary testing, or report authorisation. Finally, no external validation, calibration analysis, diagnostic-error severity grading, or cost evaluation was performed. Multiple exploratory comparisons were conducted without multiplicity correction; individual p-values should therefore be interpreted cautiously.

CONCLUSION

AI-assisted histopathological interpretation achieved higher exact concordance with the integrated reference diagnosis and substantially shorter review time than conventional interpretation alone in this 50-case cohort. All conventional discordances were corrected after AI-supported review. These results indicate potential value for AI as a diagnostic adjunct, but they do not establish autonomous diagnostic performance or routine generalisability. The perfect concordance observed in the AI arms, possible reference-standard dependence, heterogeneous case mix, and incomplete technical documentation require cautious interpretation. Larger multicentre studies with prespecified endpoints, independently blinded reference adjudication, reader-level data, external validation, calibration assessment, and clearly documented scanner and system specifications are necessary before implementation beyond a supervised decision-support role.

Source of Funding
Nil.

Conflict of Interest
None declared.

REFERENCES

1. Mukhopadhyay S, Feldman MD, Abels E, Ashfaq R, Beltaifa S, Cacciabeve NG, et al. Whole slide imaging versus microscopy for primary diagnosis in surgical pathology: a multicenter blinded randomized noninferiority study of 1992 cases. Am J Surg Pathol. 2018;42(1):39–52. doi:10.1097/PAS.0000000000000948.
2. Borowsky AD, Glassy EF, Wallace WD, Kallichanda N, Behling CA, Miller DV, et al. Digital whole slide imaging compared with light microscopy for primary diagnosis in surgical pathology: a multicenter, double-blinded, randomized study of 2045 cases. Arch Pathol Lab Med. 2020;144(10):1245–1253. doi:10.5858/arpa.2019-0569-OA.
3. Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat Med. 2019;25(8):1301–1309. doi:10.1038/s41591-019-0508-1.
4. Ehteshami Bejnordi B, Veta M, van Diest PJ, van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA. 2017;318(22):2199–2210. doi:10.1001/jama.2017.14585.
5. Bulten W, Pinckaers H, van Boven H, Vink R, de Bel T, van Ginneken B, et al. Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. Lancet Oncol. 2020;21(2):233–241. doi:10.1016/S1470-2045(19)30739-9.
6. Iizuka O, Kanavati F, Kato K, Rambeau M, Arihiro K, Tsuneki M. Deep learning models for histopathological classification of gastric and colonic epithelial tumours. Sci Rep. 2020;10(1):1504. doi:10.1038/s41598-020-58467-9.
7. Steiner DF, MacDonald R, Liu Y, Truszkowski P, Hipp JD, Gammage C, et al. Impact of deep learning assistance on the histopathologic review of lymph nodes for metastatic breast cancer. Am J Surg Pathol. 2018;42(12):1636–1646. doi:10.1097/PAS.0000000000001151.
8. Retamero JA, Gulturk E, Bozkurt A, Liu S, Gorgan M, Moral L, et al. Artificial intelligence helps pathologists increase diagnostic accuracy and efficiency in the detection of breast cancer lymph node metastases. Am J Surg Pathol. 2024;48(7):846–854. doi:10.1097/PAS.0000000000002248.
9. Pantanowitz L, Quiroga-Garza GM, Bien L, Heled R, Laifenfeld D, Linhart C, et al. An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images of core needle biopsies: a blinded clinical validation and deployment study. Lancet Digit Health. 2020;2(8):e407–e416. doi:10.1016/S2589-7500(20)30159-X.

10. Raciti P, Sue J, Ceballos R, Godrich R, Kunz JD, Kapur S, et al. Novel artificial intelligence system increases the detection of prostate cancer in whole slide images of core needle biopsies. Mod Pathol. 2020;33(10):2058–2066. doi:10.1038/s41379-020-0551-y.
11. Mundhada A, Sundaram S, Swaminathan R, D'Cruze L, Govindarajan S, Makaram N. Differentiation of urothelial carcinoma in histopathology images using deep learning and visualization. J Pathol Inform. 2023;14:100155. doi:10.1016/j.jpi.2022.100155.
12. Zehra T, Jaffar N, Shams M, Chundriger Q, Ahmed A, Anum F, et al. Use of a novel deep learning open-source model for quantification of Ki-67 in breast cancer patients in Pakistan: a comparative study between the manual and automated methods. Diagnostics (Basel). 2023;13(19):3105. doi:10.3390/diagnostics13193105.
13. Schömig-Markiefka B, Pryalukhin A, Hulla W, Bychkov A, Fukuoka J, Madabhushi A, et al. Quality control stress test for deep learning-based diagnostic model in digital pathology. Mod Pathol. 2021;34(12):2098–2108. doi:10.1038/s41379-021-00859-x.
14. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. Nat Med. 2018;24(10):1559–1567. doi:10.1038/s41591-018-0177-5.
15. Cruz-Roa A, Gilmore H, Basavanhally A, Feldman M, Ganesan S, Shih NNC, et al. Accurate and reproducible invasive breast cancer detection in whole-slide images: a deep learning approach for quantifying tumor extent. Sci Rep. 2017;7:46450. doi:10.1038/srep46450.