

**A Comprehensive Analysis of Accuracy of Gemini AI Assistance in Diagnosis of Histo-Pathological Slides**<sup>1</sup>Dr. Shaista Choudhary, Professor, Department of Pathology, Dr.B.R. Ambedkar Medical College<sup>2</sup>Dharshini S.S., 4th year MBBS student, Dr. B.R. Ambedkar Medical College<sup>3</sup>Maanas Varun Mahesh, 4th year MBBS student, Dr. B.R. Ambedkar Medical College<sup>4</sup>Amogha G. Shetty, 4th year MBBS student, Dr. B.R. Ambedkar Medical College**Corresponding Author:** Maanas Varun Mahesh, 4th year MBBS student, Dr. B.R. Ambedkar Medical College.**How to citation this article:** Dr. Shaista Choudhary, Dharshini S.S., Maanas Varun Mahesh, Amogha G. Shetty, “A Comprehensive Analysis of Accuracy of Gemini AI Assistance in Diagnosis of Histo-Pathological Slides”, IJMACR–September– 2026, Volume – 9, Issue – 5, P. No. 782 – 789.**Open Access Article:** © 2026 Dr. Shaista Choudhary, et al. This is an open access journal and article distributed under the terms of the creative common’s attribution license (<http://creativecommons.org/licenses/by/4.0>). Which allows others to remix, tweak, and build upon the work non - commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.**Type of Publication:** Original Research Article**Conflicts of Interest:** Nil**Abstract****Aim:** This study was undertaken with the objective of evaluating whether Gemini AI can function as a helping aid for pathologists by assessing its diagnostic accuracy when interpreting histopathological slides.**Methods:** Digitized histopathology slides encompassing inflammatory, benign, and malignant conditions across multiple organ systems were analyzed independently by Gemini AI and experienced pathologists, with the latter serving as the diagnostic gold standard. The degree of concordance between AI-generated interpretations and expert diagnoses was systematically assessed.**Results:** The results demonstrate that Gemini AI performs well in identifying classical and well-defined histomorphological patterns, while showing limitations in diagnostically complex or borderline lesions. These findings reinforce the concept that AI cannot replace the nuanced judgment, clinical correlation, and experience of

a pathologist. Instead, AI shows promise as an adjunctive tool that may assist in preliminary screening, workload reduction, and educational support.

**Conclusion:** In conclusion, Gemini AI shows promising diagnostic capability in histopathology, with potential applications in workload reduction, preliminary screening, and educational support. However, integration into routine clinical practice must be approached cautiously, with continued validation, ethical oversight, and pathologist supervision.**KEYWORDS:** Artificial intelligence, Gemini AI, Histopathology, Pathologists, Diagnostic support, Digital pathology**Introduction**

Histopathology forms the backbone of modern medical diagnosis, providing definitive evidence for a wide spectrum of diseases ranging from inflammatory and infectious conditions to benign and malignant

neoplasms<sup>1</sup>. The responsibility of interpreting histopathological slides rests on trained pathologists, whose expertise, pattern recognition skills, and clinical judgment are central to accurate diagnosis and patient management. Despite rapid technological advances, the role of the pathologist remains indispensable and cannot be replaced by automated systems.

In routine clinical practice, pathologists are required to examine a substantial number of slides daily. On average, a practicing surgical pathologist may review 40–80 cases per day, with each case often comprising 2–10 slides, translating to 100–400 slides per day in high-volume centers<sup>2,3</sup>. In tertiary care hospitals and cancer referral centers, this number may be even higher. The increasing burden of non-communicable diseases, aging populations, and widespread screening programs has further amplified the workload on pathology departments worldwide<sup>4</sup>. Compounding this challenge is the global shortage of trained pathologists. Several countries, including India, face a disproportionately high patient-to-pathologist ratio, leading to increased turnaround times and diagnostic fatigue<sup>5</sup>. Fatigue and time pressure are well-recognized contributors to diagnostic variability, especially in borderline lesions or cases requiring extensive slide review<sup>6</sup>. These challenges highlight the need for supportive tools that can assist pathologists without undermining their central role. Digital pathology has been a pivotal development in this direction, enabling the conversion of glass slides into high-resolution digital images suitable for remote viewing, storage, and computational analysis<sup>7</sup>. The digitization of slides has opened the door to the application of artificial intelligence (AI) in histopathology. Importantly, the purpose of introducing AI into pathology is not to replace pathologists, but to explore how technology can support

them in managing increasing workloads while maintaining diagnostic quality.

AI-based image analysis systems, particularly those using deep learning, have demonstrated the ability to recognize complex visual patterns within histopathological images<sup>8</sup>. Studies have shown that AI can assist in detecting malignancies, identifying mitotic figures, grading tumors, and highlighting areas of diagnostic interest<sup>9–11</sup>. Such capabilities suggest that AI may function as a second reader, flagging suspicious areas or providing preliminary interpretations that the pathologist can then critically evaluate. Large multimodal AI systems such as Gemini AI represent a newer generation of tools capable of integrating visual data with contextual medical knowledge<sup>12</sup>. Unlike narrow task-specific algorithms, Gemini AI is designed to generate structured diagnostic reasoning, making it particularly relevant for exploratory assessment in histopathology. When used appropriately, such systems may help pathologists triage cases, prioritize complex slides, and reduce time spent on straightforward or repetitive diagnoses<sup>13</sup>.

However, it is essential to emphasize that AI lacks the ability to perform clinical correlation, appreciate patient history, or exercise ethical and professional judgment — all of which are fundamental aspects of pathology practice<sup>14</sup>. Furthermore, histopathological diagnosis often requires subtle interpretation of architectural patterns, stromal responses, and clinicopathological correlation that extends beyond image recognition alone<sup>15</sup>. These limitations reinforce the view that AI should remain a supportive adjunct, operating under the guidance and supervision of pathologists. From a patient care perspective, the thoughtful integration of AI into pathology workflows holds significant promise. By assisting with workload management, AI may allow

pathologists to devote more time to complex cases, multidisciplinary discussions, and quality assurance activities<sup>6</sup>. This collaborative model — where AI enhances efficiency while the pathologist retains full diagnostic authority — has the potential to improve turnaround times, reduce burnout, and enhance diagnostic consistency, ultimately benefiting patient outcomes<sup>16</sup>.

The present study was conducted with this philosophy in mind. Rather than positioning Gemini AI as an alternative to human expertise, this study aims to evaluate whether Gemini AI can function as a helpful diagnostic aid in histopathological slide interpretation. By comparing AI-generated interpretations with diagnoses made by experienced pathologists across a diverse range of pathological conditions, this study seeks to identify areas where AI performs reliably, as well as scenarios where human expertise remains paramount. Such evidence is essential for guiding the responsible, ethical, and patient-centered use of AI in the future of pathology practice.

## Materials and Methods

### Study Design

This study was designed as an observational, cross-sectional analytical study aimed at assessing the diagnostic performance of Gemini AI in histopathological slide interpretation. The study did not involve human or animal experimentation and was conducted using archived histopathology slides and anonymized digital images, as documented in the institutional research application. Ethical clearance was obtained as per institutional requirements, and no patient-identifiable information was used at any stage of the study. The primary objective was to evaluate whether Gemini AI could function as a supportive diagnostic aid by comparing its interpretations with diagnoses provided by

qualified histopathologists, who were considered the gold standard.

### Study Material

The study material consisted of digitized histopathological slides obtained from the Department of Pathology, Dr. B.R. Ambedkar Medical College and Hospital. The dataset included a broad spectrum of pathological entities encompassing inflammatory, benign, premalignant, and malignant conditions across multiple organ systems. A total of 100 histopathological cases were included.

### Pathology Slide Analysis

Each case consisted of one or more hematoxylin and eosin (H&E)-stained slides representing routine diagnostic specimens. The pathological conditions evaluated included, but were not limited to:

- Hematolymphoid malignancies
- Carcinomas of breast, colon, stomach, and oral cavity
- Benign neoplasms such as fibroadenoma, pleomorphic adenoma, and multinodular goiter
- Inflammatory conditions such as appendicitis, cholecystitis, tonsillitis, and Crohn's disease
- Cystic lesions including ovarian cysts, hydrocele sac, and sebaceous cysts

This heterogeneity was intentionally maintained to reflect real-world pathology practice, rather than focusing on a single disease category.

### Reference Standard (Pathologist Diagnosis)

Each histopathological slide was independently examined by experienced histopathologists using routine light microscopy. Diagnoses provided by the histopathologists were recorded as the final reference diagnosis and served as the benchmark for comparison.

The pathologists' diagnoses were based on:

Histomorphological features

- Architectural patterns
- Cellular morphology
- Presence or absence of inflammatory, dysplastic, or malignant features

These diagnoses were documented in the pathology slide analysis records and scoring tables

Pathology Slide Analysis: No modifications were made to the original diagnoses for the purpose of this study.

### AI Tool and Prompting Methodology

Gemini AI, a large multimodal artificial intelligence model capable of image interpretation and contextual reasoning, was used for slide analysis. Digitized images of histopathological slides were provided to Gemini AI without any accompanying clinical history, laboratory findings, or prior diagnoses. A standardized prompt was used for all cases to minimize bias and ensure uniformity of responses. The prompt was:

“This is a slide of \_\_. Think like a pathologist and identify the condition.”

Gemini AI was allowed to generate a descriptive interpretation and a probable diagnosis based solely on

the visual information available. No corrective feedback or iterative prompting was provided during the analysis to avoid influencing subsequent outputs.

### Grading and Scoring System

To objectively assess the diagnostic performance of Gemini AI, a predefined grading system was employed. This grading framework was developed prior to analysis. The grading scale ranged from 0 to 5, defined as follows:

Grade 0: Unable to identify the slide

Grade 1: Unable to provide a diagnosis but able to detect an abnormality

Grade 2: Diagnosis provided with partial similarity to the actual condition

Grade 3: Accurate diagnosis in 1 out of 3 attempts

Grade 4: Accurate diagnosis in 2 out of 3 attempts

Grade 5: Accurate diagnosis in all 3 attempts

Each AI-generated diagnosis was compared with the corresponding histopathologist diagnosis and assigned an appropriate grade based on the level of concordance.

| Gemini's response                                                     | Grade |
|-----------------------------------------------------------------------|-------|
| Unable to identify the slide                                          | 0     |
| Unable to provide a diagnosis but is able to detect an abnormality    | 1     |
| Provides a diagnosis which has similarities with the actual condition | 2     |
| Provides an accurate diagnosis in 1 of 3 drafts                       | 3     |
| Provides an accurate diagnosis in 2 of 3 drafts                       | 4     |
| Provides an accurate diagnosis in 3 of 3 drafts                       | 5     |

### Data Collection and Analysis

For each case, the following data were recorded:

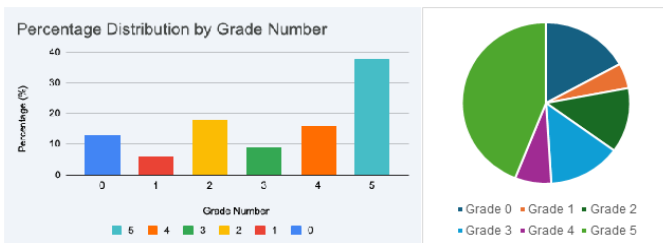
- Histopathologist diagnosis
- Gemini AI-generated diagnosis
- Assigned diagnostic grade

Data were tabulated and analyzed descriptively. Diagnostic performance was assessed by evaluating:

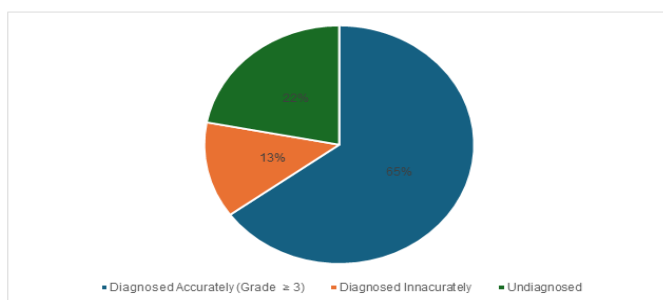
- Frequency distribution of grades
- Proportion of cases with high concordance (Grades 4 and 5) and low or non-diagnostic concordance (Grades 0–2)

## Results

The effectiveness of Gemini AI in diagnosing histopathological slides was systematically evaluated using a tailored grading system ranging from 0 to 5, where higher grades correspond to progressively greater diagnostic accuracy. This study encompassed a total of 100 slides representing a diverse spectrum of pathological conditions, allowing for a comprehensive assessment of the AI's diagnostic performance.



Graph 1



Graph 2

The grading distribution revealed distinct patterns in Gemini AI's diagnostic capability. Specifically, 17 slides (17%) were assigned Grade 0, indicating that Gemini AI was unable to identify these slides. Slides receiving Grade 1 accounted for 5% of the total, where Gemini AI detected an abnormality but failed to provide a definitive diagnosis. This suggests that while the AI can recognize

deviations from normal morphology, its diagnostic specificity remains limited in some instances. Grade 2 was assigned to 13 slides (13%), where the AI provided a diagnosis bearing similarities to the actual pathological condition. This intermediate level of accuracy indicates that Gemini AI can approximate diagnoses but may lack precision or confidence in distinguishing closely related conditions. At Grade 3, 14 slides (14%) were accurately diagnosed in at least one of the three drafts generated by Gemini AI. This demonstrates a moderate level of diagnostic reliability, where the AI can correctly identify pathology in some iterations but not consistently across all outputs. Grade 4 was awarded to 7 slides (7%), where Gemini AI provided an accurate diagnosis in two out of three drafts. These results reflect a strong diagnostic capability with near-consistent accuracy. The highest grade, Grade 5, was achieved by 44 slides (44%), where Gemini AI consistently provided accurate diagnoses across all three drafts. The substantial proportion of slides reaching Grade 5 underscores Gemini AI's robust performance in identifying common and well-characterized pathological entities.

Overall, the data indicate that Gemini AI was able to provide an accurate diagnosis (Grade 3 or higher) in approximately 65% of the cases evaluated. Within this subset, 44% of slides were diagnosed with the highest degree of accuracy (Grade 5), reflecting excellent diagnostic consistency. Conversely, the AI failed to identify or provide a meaningful diagnosis (Grades 0 and 1) in 22% of cases, highlighting limitations in recognizing certain complex or less common pathologies.

## Discussion

The present study evaluated the diagnostic performance of Gemini AI in histopathological slide interpretation across a diverse spectrum of inflammatory, benign,

pre-malignant, and malignant lesions. The inclusion of specimens from multiple organ systems enhances the external validity of the findings and reflects real-world diagnostic exposure. High diagnostic concordance was observed in morphologically classical entities such as acute appendicitis, chronic cholecystitis, multinodular goitre, pleomorphic adenoma, ovarian cysts, endometrial hyperplasia, benign prostatic hyperplasia, fatty liver, atheromatous plaque, and carcinoma breast. These lesions demonstrate well-defined architectural and cytological features, supporting the premise that artificial intelligence systems perform optimally in pattern-recognition–dominant tasks. Prior work in computational pathology has similarly shown strong AI performance in well-characterised malignant and benign lesions with distinct histomorphology<sup>21</sup>. Reduced concordance was noted in diagnostically complex or context-dependent lesions, including reactive lymphadenopathy, granulomatous inflammation, inflammatory myofibroblastic tumour, and certain early malignant lesions. Such cases often require clinicopathological correlation and immunohistochemical (IHC) evaluation. Generative multimodal AI systems primarily rely on learned pattern associations rather than mechanistic reasoning, which may limit their performance in subtle or overlapping pathological entities<sup>22</sup>.

Systematic reviews and meta-analyses of artificial intelligence in digital pathology report pooled sensitivities and specificities frequently exceeding 85–90% for common malignancies such as breast and colorectal carcinoma<sup>23</sup>. However, performance variability increases in inflammatory, rare, or borderline lesions — a trend mirrored in the present study. This concordance with existing literature strengthens the validity of the findings. Comparative analyses between generative AI

systems and physicians indicate that while AI may achieve comparable performance in structured diagnostic tasks, it remains limited in complex clinical reasoning scenarios requiring contextual integration and prioritisation of differential diagnoses<sup>24</sup>. The reduced diagnostic precision observed in reactive and granulomatous lesions in this study aligns with these findings. A notable strength of Gemini AI was its reproducible morphological descriptions, suggesting potential value as a teaching aid or preliminary screening tool. However, limitations include absence of independent integration of clinical data and ancillary investigations, inability to independently interpret IHC, and possible dataset bias due to underrepresentation of region-specific pathologies. From an ethical and governance standpoint, the World Health Organization emphasizes that artificial intelligence systems in healthcare must operate under appropriate human oversight and should augment, not replace, professional medical judgment<sup>25</sup>. The findings of this study support this position. Future research should incorporate larger multicentric cohorts, blinded validation designs, quantitative sensitivity and specificity analysis, and direct comparison with other multimodal AI platforms. Integration of clinicopathological and IHC data may further enhance diagnostic performance.

## Conclusion

In conclusion, Gemini AI demonstrates strong diagnostic concordance in classical histopathological entities but reduced reliability in complex, inflammatory conditions, and rare lesions. While AI serves as a formidable ally in enhancing diagnostic speed and as a viable educational tool, it remains a supplement rather than a replacement for the nuanced expertise of a pathologist.

## References

1. Rosai J. Rosai and Ackerman's surgical pathology. 11th ed. Philadelphia: Elsevier; 2018.
2. Elmore JG, Longton GM, Carney PA, Geller BM, Onega T, Tosteson AN, et al. Diagnostic concordance among pathologists interpreting breast biopsy specimens. *JAMA*. 2015;313(11):1122–32.
3. Williams BJ, Lee J, Oien KA, Treanor D. Digital pathology access and usage in the UK: results from a national survey. *J Clin Pathol*. 2018;71(5):463–6.
4. Metter DM, Colgan TJ, Leung ST, Timmons CF, Park JY. Trends in the US and Canadian pathologist workforce from 2007 to 2017. *Arch Pathol Lab Med*. 2019;143(5):536–45.
5. Fleming KA, Naidoo M, Wilson M, Flanagan J, Horton S, Kuti MA, et al. An essential pathology package for low- and middle-income countries. *Am J Clin Pathol*. 2017;147(1):15–32.
6. Zafar A, Kaur H, Sood N. Workload, stress, and burnout among pathologists: a growing concern. *Indian J Pathol Microbiol*. 2020;63(3):353–8.
7. Pantanowitz L, Sinard JH, Henricks WH, Fatheree LA, Carter AB, Contis L, et al. Validating whole slide imaging for diagnostic purposes in pathology. *Arch Pathol Lab Med*. 2013;137(12):1710–22.
8. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60–88.
9. Campanella G, Hanna MG, Geneslaw L, Miraflor A, Silva VWK, Busam KJ, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med*. 2019;25(8):1301–9.
10. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in breast cancer. *JAMA*. 2017;318(22):2199–210.
11. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyo D, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med*. 2018;24(10):1559–67.
12. Google DeepMind. Gemini: a family of multimodal artificial intelligence models. Technical report. 2023.
13. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25(1):44–56.
14. Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. *BMJ*. 2019;365:l737.
15. Fletcher CDM. Diagnostic histopathology of tumors. 5th ed. Philadelphia: Elsevier; 2021.
16. Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: challenges and opportunities. *Med Image Anal*. 2016;33:170–5.
17. Oakden-Rayner L. Exploring large-scale medical image datasets. *Lancet Digit Health*. 2020;2(7):e331–2.
18. Zormpas-Petridis K, Jørgensen KJ, Løkkegaard T, Nielsen PS. Artificial intelligence in pathology education: a systematic review. *Histopathology*. 2020;76(5):703–10.
19. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. *JAMA*. 2018;320(21):2199–200.
20. World Health Organization. Global strategy on digital health 2020–2025. Geneva: WHO; 2021.

21. Campanella G, Hanna MG, Geneslaw L, 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
22. Yang L, Xu S, Sellergren A, et al. Advancing biomedical understanding with multimodal Gemini. *arXiv*. 2024; 2401.XXXX.
23. Steiner DF, MacDonald R, Liu Y, 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
24. Blease C, Bernstein MH, Gaab J, et al. Comparing clinician and generative AI diagnostic reasoning: a systematic review and meta-analysis. *NPJ Digit Med*. 2025;8: XX.
25. World Health Organization. Ethics and governance of artificial intelligence for health. Geneva: WHO; 2021.