



# Oral Maxillofacial Pathology Specialist and AI Supported for Histopathological Diagnosis of Oral Lesions

Ilham Nur Muhammad<sup>1\*</sup>, Reza Al Fessi<sup>1</sup>, Sisca Meida Wati<sup>2</sup>

<sup>1</sup>Department of Oral and Maxillofacial Surgery, Faculty of Dental Medicine, Universitas Airlangga, Surabaya, Indonesia

<sup>2</sup>Department of Oral and Maxillofacial Pathology, Faculty of Dental Medicine Universitas Airlangga, Surabaya, Indonesia

\* Corresponding author.

E-mail address: [inm.ilhamnur@gmail.com](mailto:inm.ilhamnur@gmail.com)

---

## Article information

Submitted  
13-01-2026

Accepted  
21-02-2026

Published  
28-02-2026

## Abstract

**Background:** Histopathological (HPA) analysis is the gold standard for oral lesions. However, clinical diagnoses often have low agreement rates (44.1%). Artificial intelligence (AI) may offer diagnostic assistance, but its accuracy requires critical evaluation. Purpose: This study compared the diagnostic accuracy of AI models, ChatGPT and Gemini, in interpreting HPA images of oral maxillofacial lesions to the gold standard diagnosis from Oral and Maxillofacial Pathology (OMP) specialists.

**Methods:** This analytical observational study used 54 digital HPA slides from a university research center. The diagnoses generated by ChatGPT and Gemini were evaluated for agreement ('Correct' or 'Incorrect') with the definitive diagnoses made by OMP specialists. Ethical approval was obtained.

**Results:** Gemini demonstrated a diagnostic accuracy of 74.07% (40 of 54 cases), while ChatGPT achieved 70.37% (38 of 54 cases). The most common lesions were mucoceles and dentigerous cysts. A statistically significant difference (<0.001) was observed between the accuracy of both AI models and the OMP specialist.

**Conclusion:** AI models showed considerable ability to recognize histopathological patterns, but their accuracy was significantly lower than OMP specialists. AI is an augmentative tool for triage or learning but cannot replace the role of OMP specialists in establishing a definitive diagnosis.

**Keywords:** Artificial Intelligence, Oral and Maxillofacial Lesions, Histopathology, Diagnostic Accuracy

---

## Introduction

Histopathological examination is essential in the diagnosis of oral lesions because many lesions present with similar clinical features. A definitive diagnosis is achieved only through microscopic evaluation of tissue specimens, which guides appropriate management strategies and assists in predicting prognosis. The interpretation provided by oral and maxillofacial pathologists represents the global gold standard for diagnosing oral lesions due to its precision in assessing cellular and tissue alterations.<sup>1</sup>

The spectrum of oral and maxillofacial lesions predominantly includes cystic, inflammatory, and reactive entities, with the mandible and maxilla serving as the most frequent biopsy sites. Previous studies reported that the clinical–histopathological concordance rate was approximately 44%, highlighting the limitations of clinical assessment alone and emphasizing the need for microscopic confirmation.<sup>2</sup> Consequently, histopathological examination performed by specialized pathologists remained the gold standard, offering superior diagnostic accuracy through detailed evaluation of tissue architecture and cellular morphology to differentiate benign from malignant conditions. This diagnostic dependence underscored the importance of interdisciplinary collaboration between oral surgeons and pathologists in managing maxillofacial lesions.<sup>3</sup>

The rapid development of artificial intelligence (AI) offers new opportunities to support clinical decision-making and improve diagnostic efficiency. Several studies demonstrated that AI-assisted diagnostic systems improved accuracy in image-based medical analysis.<sup>4,5</sup> In oral pathology, current advancements predominantly utilized task-specific Convolutional Neural Networks (CNNs) trained on curated histopathological datasets. Although CNN-based systems showed high diagnostic performance, their clinical applicability was often limited by narrow domain specificity. The emergence of Large Multimodal Models (LMMs), such as ChatGPT and Gemini, introduces a general-purpose AI paradigm capable of integrating both textual and visual data simultaneously.<sup>6,7</sup>

This study evaluates the viability of ChatGPT and Gemini as artificial intelligence–based clinical decision support tools for early-stage screening in oral pathology. In peripheral clinical settings where access to specialized oral and maxillofacial pathologists is limited, AI-assisted diagnostic support may help reduce delays in diagnosis and alleviate pressure on healthcare systems.

## Methods

This analytical observational study employed a descriptive comparative design and evaluated 54 digital histopathological slides (JPEG/PNG format; resolution  $\geq 1080p$ ) retrieved from the archives of the Oral and Maxillofacial Pathology Laboratory, Faculty of Dental Medicine, Universitas Airlangga. The study was conducted under ethical clearance number KKEPK.FKG/X/2025. Each slide had an established definitive diagnosis made by experienced oral and maxillofacial pathologists, which served as the gold standard reference. All data were anonymized to maintain patient confidentiality.

To minimize bias, both the oral pathologists and the artificial intelligence (AI) models were blinded to the definitive clinical diagnoses and to each other's assessments. All evaluations were performed independently and were based solely on standardized diagnostic prompts and the histopathological images provided.

The AI models were assessed using a standardized multimodal prompting strategy. For each case, relevant clinical information was integrated with high-resolution histopathological images. The models were instructed to evaluate five diagnostic domains: cellular architecture, cellular morphology, nuclear chromatin characteristics, stromal and inflammatory infiltration, and final diagnosis. Each sample underwent a single independent evaluation cycle to measure immediate diagnostic performance and the ability to correlate clinical and histopathological findings.

Diagnostic agreement between the AI models and the expert pathologists was analyzed. A diagnosis was categorized as “correct” when the AI output matched the expert diagnosis and as “incorrect” when discrepancies occurred either clinically or histologically. Diagnostic accuracy was calculated as the proportion of correct classifications. Differences in diagnostic accuracy between the AI models and the oral and maxillofacial pathology (OMP) specialist were analyzed using the Wilcoxon signed-rank test, with statistical significance determined based on the corresponding p-values.

**Disclosure of AI Use:** The authors used the AI tools ChatGPT (OpenAI, 2025 version) and Gemini (Google DeepMind, 2025 version) during the study for multimodal image interpretation as described above. All AI-generated outputs were critically reviewed and edited by the authors, who assumed full responsibility for the final content of this publication.

## Results

A total of 54 cases were analyzed. The case distribution demonstrated a diverse spectrum of lesion types. Mucocele and Dentigerous Cyst were the most frequently identified lesions, each accounting for 10 cases (18.52%). Fibromatous Epulis, Ameloblastoma, and Irritation Fibroma were also commonly observed, with 6 cases each (11.11%). Radicular Cyst was identified in 4 cases (7.41%). Leukoplakia, Lipoma, Odontogenic Keratocyst (OKC), Nasopalatine Duct Cyst, Fissuratum Epulis, and Granulomatous Epulis represented the least frequent diagnoses, with 2 cases each (3.70%).

Distribution of oral lesion cases at the Research Center of the Faculty of Dentistry, Airlangga University, January-August 2025 (Tabel 1).

**Table 1.** Distribution of oral lesion cases at the Research Center of the Faculty of Dentistry, Airlangga University, January-August 2025

Oral Lesions Diagnosis	Number of Cases Slide HPA	Percentage (%)
Leukoplakia	2	3.70
Lipoma	2	3.70
OKC	2	3.70
Mucocele	10	18.52
Epulis Fibromatosa	6	11.11
Nasopalatinus Duct Cyst	2	3.70
Radicular Cyst	4	7.41
Ameloblastoma	6	11.11
Dentigerous Cyst	10	18.52
Iritasi Fibroma	6	11.11
Epulis Fissuratum	2	3.70
Epulis Granulomatosa	2	3.70
<b>Total Case</b>	<b>54</b>	<b>100</b>

The case distribution indicated that the majority of analyzed cases were benign and reactive lesions, predominantly consisting of cystic lesions and non-neoplastic soft tissue lesions. This distribution aligned with epidemiological patterns of oral lesions previously reported at other research centers in Indonesia, where mucoceles and dentigerous cysts frequently resulted from chronic inflammation or minor salivary gland obstruction.

The presence of ameloblastoma and odontogenic keratocyst (OKC) cases also reflected a considerable proportion of odontogenic lesions, which required accurate histopathological assessment to distinguish between benign and more aggressive biological behavior. Overall, these findings underscored the importance of thorough histopathological evaluation in establishing definitive diagnoses within oral and maxillofacial pathology.

Diagnostic Accuracy between GPT and Gemini AI (Tabel 2).

**Table 2.** Diagnostic Accuracy between GPT and Gemini AI

	Diagnostic Accuracy						Correct Diagnosis Percentage	Cohens Kappa Coeff.
	Cellular Architecture	Cellular Morphology	Nuclear Chromatin	Stroma	Inflammatory Cells	Diagnosis		
GPT	38	36	34	38	42	38	70.37	<0.001*
Gemini	40	34	34	40	40	40	74.07	0.000*

The diagnostic accuracy of the evaluation of 54 histopathology slides of oral and maxillofacial lesions varied among evaluator groups. Oral and Maxillofacial Pathologists (OMP) achieved 100% diagnostic accuracy, which served as the gold standard reference. The ChatGPT model generated correct diagnoses in 38 of 54 cases (70.37%), whereas the Gemini model produced 40 correct diagnoses (74.07%).

Both generative AI architectures demonstrated substantial proficiency in identifying histomorphological patterns across a diverse cohort of oral pathologies. Gemini exhibited a marginal performance advantage over ChatGPT; however, both models showed a diagnostic deficit of approximately 25–30% relative to the 100% accuracy achieved by specialized clinicians. Cohen's Kappa coefficient analysis revealed a statistically significant discrepancy between the AI-generated diagnoses and the interpretations provided by oral and maxillofacial pathology specialists as the gold standard. The comparative agreement analysis further indicated that Gemini and ChatGPT demonstrated relatively similar diagnostic capabilities.

## Discussion

Clinical diagnosis without cytological or histopathological examination showed lower accuracy and depended on clinician experience and lesion characteristics.<sup>2</sup> Based on the evaluation results, Artificial Intelligence (AI) models, including GPT and Gemini, achieved diagnostic accuracies of 70.37% and 74.07%, respectively. Although these percentages reflected substantial diagnostic capability, the results remained significantly different from the 100% accuracy achieved by oral and maxillofacial pathology specialists, who served as the gold standard for histopathological diagnosis.<sup>1</sup>

The integration of Artificial Intelligence within oral and maxillofacial pathology demonstrated considerable potential in improving diagnostic efficiency and streamlining workflow, particularly in lesion classification and high-risk case detection. However, important limitations persisted in managing complex diagnostic scenarios that required nuanced clinical judgment.<sup>10,11</sup>

In this study, although AI models demonstrated commendable performance, their accuracy remained statistically inferior to that of Oral and Maxillofacial Pathology (OMP) specialists ( $p=0.000$ ). These findings were consistent with previous reports that documented diagnostic accuracies ranging from 68–80% depending on morphological complexity, and studies that identified challenges in AI interpretation of chronic inflammatory infiltrates and tissue artifacts.<sup>12,13</sup> The observed discrepancy highlighted a fundamental limitation of current generative models, namely their reliance on isolated visual pattern recognition rather than comprehensive integration of clinical, radiographic, and pathological data that characterized human expertise.<sup>14</sup>

Although AI performed effectively in high-throughput screening and in detecting overt malignant features, its reliability decreased when differentiating complex premalignant lesions or reactive entities that mimicked neoplastic processes. In addition, inherent dataset biases toward common pathologies reduced sensitivity in detecting rare odontogenic tumors or atypical variants. Imbalances in training datasets further contributed to diagnostic variability.<sup>6</sup>

Consequently, models such as ChatGPT and Gemini functioned as valuable adjunctive tools for triage and digital pathology education; however, they did not replace the multidimensional interpretive expertise of specialized pathologists, particularly in borderline or atypical cases requiring contextual synthesis.<sup>16–18</sup>

The integration of Artificial Intelligence in digital pathology significantly reduced processing time for preliminary slide evaluation.<sup>19</sup> Nevertheless, diagnostic precision remained the primary objective in patient management and required validation by specialized professionals to minimize the risk of algorithmic “black-box” errors.<sup>20</sup> Beyond clinical application, AI also functioned as an educational tool that supported accelerated learning among residents, particularly in the context of the global shortage of Oral and Maxillofacial Pathology specialists.<sup>9,21</sup> Overall, AI operated as an augmentative system that supported, rather than replaced, definitive anatomopathological diagnosis.<sup>15,22–24</sup>

This study represented a comprehensive domain-specific analysis that systematically evaluated AI performance from cellular architecture assessment to definitive diagnosis under a strict double-blinded protocol. This design ensured objective comparison between general-purpose AI systems and expert pathologists while minimizing observer bias. However, limitations included dataset homogeneity, which might not have represented the full spectrum of rare maxillofacial variants. In addition, undisclosed model updates could have influenced diagnostic reproducibility. Future research will integrate larger multicenter datasets and will investigate explainable AI frameworks to enhance clinical transparency and reliability.

## Conclusions

This study demonstrated that Large Multimodal Models (LMMs), specifically ChatGPT and Gemini, exhibited significant potential as clinical decision support tools in oral pathology by providing detailed interpretations of histopathological features. Although both models showed proficiency in identifying cellular architecture and morphology, their diagnostic accuracy remained inferior to the expert histopathological gold standard.

The integration of AI-driven risk stratification improved the potential efficiency of the triage-to-referral process, particularly in resource-constrained peripheral settings. However, further validation using Whole Slide Imaging (WSI) and multi-iteration reliability protocols was required to ensure diagnostic safety and longitudinal reproducibility before these general-purpose models were integrated into routine clinical workflows.

## Acknowledgements

The authors would like to express their gratitude to Research Center Faculty of Dental Medicine Universitas Airlangga for providing the facilities and resources necessary to conduct this research.

## Declaration concerning generative AI and AI-augmented technologies in the compositional process

In the course of preparing this paper, the authors utilized ChatGPT to enhance readability and linguistic quality. Subsequent to utilizing this tool/service, the writers assessed and amended the information as necessary and assume complete accountability for the publication's content.

## Declarations of competing interest

No potential competing interest was reported by the authors.

## References

1. Alsanie IS, Qannam A, Bello IO, Khurram SA. Exploring the role of artificial intelligence in oral pathology: diagnostic and prognostic implications. *J Oral Pathol Med*. 2025 Aug;54(7):487-497.
2. Sindi AM, Aljohani K. Agreement between clinical and histopathological diagnoses of oral and maxillofacial lesions and influencing factors: a five-year retrospective study. *Clin Cosmet Investig Dent*. 2024 Aug 28;16:273-282.
3. Ernawati DS. Ilmu penyakit mulut (oral medicine) sebagai jembatan yang memfasilitasi ilmu. Pidato pengukuhan guru besar. Surabaya: Universitas Airlangga; 2011.
4. Tolstaya E, Tichy A, Paris S, Schw J, Aarabi G, Chaurasia A, et al. Machine learning versus clinicians for detection and classification of oral mucosal lesions. *J Dent*. 2025;161:105992.
5. Patil S, Albogami S, Hosmani J, Mujoo S, Kamil MA, Mansour MA, et al. Artificial intelligence in the diagnosis of oral diseases: applications and pitfalls. *Diagnostics (Basel)*. 2022;12(5):1029
6. Xu Z, Lin A. Current AI Applications and Challenges in Oral Pathology. 2025; 5(1):2.
7. Guler R, Yalcin E. Performance of AI chatbots in preliminary diagnosis of maxillofacial pathologies. *Med Sci Monit*. 2025;31:e949076.
8. Hegde S, Ajila V, Zhu W, Zeng C. Artificial intelligence in early diagnosis and prevention of oral cancer. *Asia Pac J Oncol Nurs*. 2022;9(12):100133.
9. Vinay V, Jodalli P, Chavan MS, Buddhikot CS, Testarelli L. Artificial intelligence in oral cancer: a comprehensive scoping review of diagnostic and prognostic applications. *Diagnostics (Basel)*. 2025;15(3):280
10. El-khoury R. The Rise of AI-Assisted Diagnosis : Will Pathologists Be Partners or Bystanders ? 2025;(1285):1–16.
11. Abdul NS, Shivakumar GC, Sangappa SB, Blasio M Di, Crimi S, Ciccì M, et al. Applications of artificial intelligence in the field of oral and maxillofacial pathology : a systematic review and meta - analysis. *BMC Oral Health*. 2024;24;(1):1–12.

12. Li X, Zhao D, Xie J, Wen H, Liu C, Li Y, et al. Deep learning for classifying the stages of periodontitis on dental images : a systematic review and meta-analysis. 2023;1–23.
13. Lee L yu, Yang C han, Lin Y chieh, Hsieh Y han, Chen Y an, Chang MD tsyr, et al. A domain knowledge enhanced yield based deep learning classi fi er identi fi es perineural invasion in oral cavity squamous cell carcinoma. 2022;12(11):1–15.
14. Khoury ZH, Ferguson A, Price JB, Sultan AS, Wang R. Responsible artificial intelligence for addressing equity in oral healthcare. 2024;1(7):1–6.
15. Khanagar SB, Al-Ehaideb A, Patil S, Baeshen HA, Sarode SC. Developments, application, and performance of artificial intelligence in dentistry: a systematic review. J Dent Sci. 2021 Jan;16(1):508-522.
16. Sriram A, Ramachandran K, Krishnamoorthy S. Artificial Intelligence in Medical Education : Transforming Learning and Practice. 2025;17(3):1-10.
17. Sriram T, B GJ. From algorithm to applications : Artificial intelligence – A future prospective in medicine. 2025;4(2):44–52.
18. Bellahsen-harrar Y, Lubrano M, Lépine C. Exploring the risks of over-reliance on AI in diagnostic pathology . What lessons can be learned to support the training of young pathologists ? 2025;20(8):1–13.
19. Gharat MG, Deshpande SM, Dhone S, Shreesh V. Digital pathology: revolutionizing oral and maxillofacial diagnostics. 2024;20(12):1834-1840.
20. Chavarkar S. Artificial intelligence in pathology: bridging the gap between technology and diagnostics. 2026;14(1):251-264.
21. Abdollahi A. The role of artificial intelligence in the future of pathology. Iran J Pathol. 2026;21(1):160-162.
22. Khalaf A, Alkhalaf A, Alshammari MM. Integration of artificial intelligence in histopathological and radiological image analysis: enhancements in diagnostic workflow. 2024;8(S1):938-953.
23. Nieri M, Serni L, Clauser T, Paoletti C, Franchi L. Diagnosis of oral cancer with deep learning: a comparative test accuracy systematic review. Oral Dis. 2025 Aug;31(8):2368-2381.
24. McGenity C, Clarke EL, Jennings C, Matthews G, Freduah-Agyemang H, Stocken DD, et al. Artificial intelligence in digital pathology: a systematic review and meta-analysis of diagnostic test accuracy. 2024;7(1):114.