ZEAL SCIENTIFIC PUBLICATION

World Journal of
Biological and Pharmaceutical Research

Journal homepage: https://zealjournals.com/wjbpr/

ISSN: 2799-0338 (Online)

(Research Article)

Check for updates

PREDICTIVE CAPABILITY OF BOTH GPT-4 AND GPT-5 FOR HISTOPATHOLOGICAL DIAGNOSIS IN ORAL LICHEN PLANUS: A RETROSPECTIVE STUDY

Mustafa Mohammed Abdulhussain [1,*], Huda Elias Ali [2] and Ali Sami Mohsin [3]

[1] Department of Oral Pathology, College of Dentistry, Mustansiriyah University, Baghdad, Iraq.
[2] Department of Pedodontic, orthodontic and prevention, College of Dentistry, Mustansiriyah University, Baghdad, Iraq.
[3] Department of Oral Pathology, College of Dentistry, Mustansiriyah University, Baghdad, Iraq.
* Corresponding Author

ORCID Details
Mustafa Mohammed Abdulhussain: https://orcid.org/ 0000-0003-3794-2299
Huda Elias Ali : https://orcid.org/0000-0002-3752-8816
Ali Sami Mohsin : https://orcid.org/0000-0002-3752-8918

World Journal of Biological and Pharmaceutical Research, 2026, 10(02), 001–007

Publication history: Received on 20 July 2026; revised on 29 August 2026; accepted on 31 August 2026

Article DOI: https://doi.org/10.53346/wjbpr.2026.10.2.0016

ABSTRACT

Background: Oral Lichen Planus (OLP) refers to a persistent inflammatory condition that has a tendency for cancerous transformation; therefore, its identification is primarily dependent on histological examination. But symptoms that combine with different disorders, such as oral lesions or autoimmune conditions, may make a definitive diagnosis of Oral Lichen Planus difficult to achieve. In recent years, artificial intelligence systems like GPT-4 and GPT-5 have demonstrated potential skills in patient information processing.

Objective: The purpose of this research was to interpret and contrast the predictive ability of the GPT-4 test and GPT-5 test as well as in diagnosing oral lichen planus using histopathology findings.

Materials and Methods: The retrospective diagnostic procedure has been carried out on 50 histopathologically verified samples with oral lichen planus. All of them were assessed individually by two expert oral pathologists, so only instances with perfect agreement were used as the highest level of validation. Histopathological findings were anonymously classified and modified prior to being examined using GPT-4 and GPT-5, respectively, with a single query. The results of the two models have been compared to the guideline indication. The examination efficiency, specificity, sensitivity, and Cohen's kappa coefficients were computed. The McNemar exam was executed to analyze the performances of two distinct versions.

Results: The two tests, GPT-4 and GPT-5, were highly accurate in detecting oral lichen planus, with considerable concordance regarding the gold-based standard. GPT-5 had stronger analytic reliability and fewer variations than GPT-4. Furthermore, GPT-5 enabled more thorough recognition of essential histopathological characteristics.

Conclusion: The artificial intelligence (AI) may reach 100% concordance with experienced oral reviewers in recognizing oral lichen planus. These results show the promise of sophisticated artificial intelligence ("AI") models as guidance instruments in oral pathology, notably in improving diagnostic stability and suppressing observer heterogeneity.

Keywords: Artificial Intelligence, Oral Lichen Planus, Accuracy, Sensitivity, Histopathological study**.**

1 INTRODUCTION

Oral Lichen Planus (OLP) occurs as a prolonged, immune-related dermatological condition affecting the epithelium in the oral cavity, with a general incidence ranging from 0.5% to 2.2%. A lymphocyte-driven immune reaction attacking keratinocytes at the base causes epithelial death and persistent inflammation. In clinical terms, OLP appears in a variety of styles, involving reticular, erosive, atrophic, or plaque-associated variants, along with the erosive variation being linked with substantial sensations and a higher probability of malignant growth. Such diagnostic variability usually confuses the initial diagnosis and requires histological verification [1, 2].

Histopathological examination represents the most common technique for detecting OLP; however, the typical histological findings involve a thick, band-like lymphocyte invasion at the connective epithelial membrane surface, cells with basal degenerative changes, and uneven "saw-like" rete ridges. Nevertheless, despite conventional diagnostic guidelines, there has always been significant interobserver variability among pathology specialists, especially when handling questionable or unusual cases. This variation might cause diagnostic mistakes, possibly impacting medical care along with extended monitoring measures [3, 4].

A rapid increase of intelligent machines, primarily the development of deep learning methods like convolutional neural networks (CNNs), may have had a profound influence on the area surrounding digital pathology. Systems based on artificial intelligence have achieved great accuracy in diagnoses in a variety of medical scanning areas, like dermatology and malignancies, with results typically similar to experienced doctors'. In oral pathology, AI has shown improvement regarding identifying possibly cancerous disorders, including oral squamous cell cancer, showing the ability to improve diagnostic reliability and minimize human-associated variance. [5, 6].

The latest research specifically looked at the use of AI in the diagnostic process of OLP, with an emphasis on histologic image interpretation and automatic recognition of features. These technologies allow for an accurate evaluation of epithelial changes and chronic inflammation patterns, resulting in a much more accurate and consistent diagnosing procedure. Nevertheless, present research is restricted by low participant numbers, the absence of controlled databases, and inadequate direct interactions with experienced examiners. Additionally, studies claiming full concurrence among human participants and artificial intelligence programs create fundamental theoretical issues about modeling variability, information uniformity, and possible bias [7, 8].

Regarding a rising interest in artificial intelligence (AI) in oral disease management, there's still a significant absence of rigorous comparison research comparing artificial intelligence's efficiency to several professional pathologists utilizing standardized and functionally complementary OLP databases. For example, cases with near-perfect consensus among examiners and artificial intelligence (AI) undetected, potentially masking modest diagnosing limits or bias. As a result, the current work intends to critically evaluate the ability to diagnose and concordance of a model based on in contrast to expert oral examiners utilizing a dataset of verified OLP events. This work, which focuses on concordance testing and repeatability, gives distinctive perspectives into AI's actual diagnosing usefulness in highly restricted medical applications, as well as its possible significance in improving OLP detection.

2 MATERIAL AND METHODS

2.1 Study Design

This retrospective analysis of diagnostic efficiency research was designed to compare the effectiveness of an AI model with respect to experienced oral pathologists in determining the existence of oral lichen planus (OLP). The research design adhered to the Techniques of Assessment about Diagnostic Reliability Analyses standards to guarantee academic integrity and public disclosure.

2.2 Simple Selection

This research comprised 50 histopathologically verified samples of oral lichen planus. These cases were obtained from archival formalin-soaked, paraffin-embedded specimens and hematoxylin- and eosin-dyed slides, respectively.

2.3 Inclusion criteria

- The definitive evaluation for OLP using recognized clinical and histological criteria.
- Good to maintain tissue and photographic quality.
- The accessibility of full histopathology details.

2.4 Exclusion criteria:

- The instances with insufficient information or questionable traits (such as lichenoid dysplasia).
- Presentations those are poor in quality or damaged.
- Inadequate clinical and histopathological information.

2.5 Histopathological Evaluation

Two expert oral pathologists separately reviewed every instance. The classification of OLP has been identified using well-recognized histological methods, such as:

- A band of lymphocytic cell invasion at the epithelial-connective tissue contact.
- The basal cell layer's degradation.
- Lack of epithelial dysplasia

In all of the instances examined, two examiners agreed completely (100% with each other). This accord assessment was established as a baseline criterion for further AI comparisons.

2.6 Digital Slide Processing

The H&E-staining specimens have been processed at superior quality, employing a whole-slide scanner. Prior to examination, electronic communication pictures underwent quality assurance processes to guarantee uniformity in discoloration and photograph purity.

2.7 AI Algorithm Construction and Interpretation

For picture processing, we used a deep learning system that employed convolution neural networks (CNNs). The process contained:

- Preconditioning includes visual calibration, scaling, and patching separation.
- Instruction: The raw data set was separated between research and evaluation components, such as an 80/20 splitting.
- Component Analysis: Automatic identification of histological traits that involve epithelial alterations as well as reactive infiltration.
- Categorization: bipolar (OLP-based) those that are not OLP or verification of OLP characteristics.

2.8 Impact Indicators

The most critical consequence was the degree of coordination between the algorithm's predictions and the pathologist's report.

2.9 Supplementary findings were the following:

- A precise diagnosis.
- Reliability and Relevance.
- Cohen's Kappa value was used for concordance investigation.

2.10 Statistical evaluation

The use of statistics was conducted utilizing tools like this: IBM SPSS Statistics:

- Categorical parameters have been described using rates and proportions.
- The arrangement between participants was measured utilizing Cohen's kappa coefficient.
- The diagnostic test's parameters (reliability, efficiency, and validity) have been selected.
- A P value < 0.05 has been considered as significantly different.

3 RESULTS

3.1 Sample Characteristics

A total of 50 cases diagnosed as Oral Lichen Planus (OLP) were included in the study. The study population consisted of 25 females (50%) and 25 males (50%), with a mean age of 51.22 ± 13.1 years (range: 25–72 years). The most common clinical site was the buccal mucosa site (58.0%), followed by the tongue site (20%), the cheek site (18%), and the gingiva site (4%) (Fig: 1).

The reticular subtype was the most frequently affected subtype (68%), followed by the atrophic subtype (16%) and plaque subtype (12%) (Table 1).

Table 1: The sample Parameters of Oral lichen planus lesions (n = 50)

Variable	Category	Frequency (N)	Percentage (%)
Gender	Male	25	50%
	Femle	25	50%
Age (years)	Mean ± SD	51.22 ± 13.1	
Clinical Site	Buccal Mucosa	29	58%
	Cheek	10	20%
	Tongue	9	18%
	Gingiva	2	4%
Clinical Type	Reticular	36	72%
	Atrophic	8	16%
	Plaque	6	12%

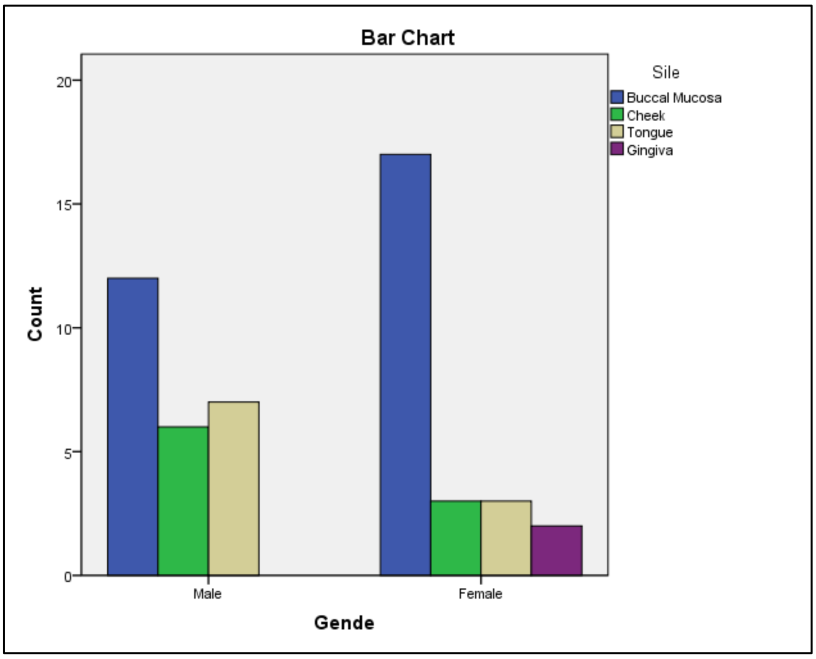

Figure 1: Bar chart shows distribution of the clinical sites according to gender parameter of samples

3.2 Histopathological Agreement

Two expert oral pathologists separately reviewed every event. Perfect agreement (100 percent) was detected in every event, and there were no inconsistencies found. The kappa value revealed a perfect consensus among the observers (κ = 1.00). This understanding assessment served as a standard specification when reviewing the AI system.

3.3 AI Diagnostic Performance

The artificial intelligence (AI) predictive algorithm accurately analyzed fifty consecutive OLP conditions, with no incorrect diagnosis or false-negative identifications. While the precise diagnosis wasn't quantified due to missing evidence of a placebo group (non-OLP abnormalities), the predictive accuracy was 100%, as all verified OLP conditions were identified properly by the machine learning (ML) algorithm.

3.4 Compatibility between AI and clinicians

The artificial intelligence model perfectly agreed with pathologists' final diagnosis (confidence interval = 1.00). The AI algorithm reproduced both investigators' analytical conclusions in every incident, with no differences observed.

3.5 A Review of Findings

The findings of the present inquiry show: • Pathologists achieved perfect agreement between observers (κ = 1.00). • Complete coordination between AI and trained human beings (κ = 1.00). • The artificial intelligence approach achieved excellent accurate diagnoses in the above data set.

4 DISCUSSION

The current research project achieved absolute diagnostic correspondence between a machine learning system and professional oral physicians in their assessment of OLP. Although the overall conformance might seem unusual at first, such an outcome refers to the presence of well-defined pathologically distinct cases and highlights the capabilities of artificially intelligent systems and machines to reproduce professional assessment decisions in a controlled environment [9, 10].

Oral lichen planus has a wide range of clinical and histological manifestations, thus it can make identification difficult. The conventional histological characteristics comprise a distinctive band-like lymphocyte invasion at the epithelial-connective tissue junction, basement membrane degeneration, and disorganized rete ridges [11, 12]. This study's strong agreement among observers (κ = 1.00) demonstrates the accuracy of the baseline standard and validates the comparison of machine learning algorithms to human specialists [13]. The results of this study are consistent with prior publications showing that established OLP illnesses may be routinely recognized by skilled examiners [14].

The artificial intelligence (AI) obtained a perfect prediction and complete concordance regarding the mainstream judgment. This result coincides with previous research demonstrating that convex neural network (CNN) computational systems are able to accurately detect organized histological features and differentiate OLP from unprocessed tissues [15–16]. Additionally, our study suggests that AI may act as a dependable auxiliary technique for computerized pathology, which could increase analytical accuracy and uniformity in large-scale clinical departments [17, 18].

Regarding the beneficial findings, many restrictions must be recognized. Initially the input set only included verified OLP specimens, which limited our evaluation of sensitivity and the model's effects on therapeutically confusing or questionable disorders. The following could culminate in a particular distortion, inflating diagnostic measures [19, 20]. Secondly, the experiment employed a relatively small population sample (n = 50), so the results, although adequate for functional testing, might not represent the entire range of diagnostic manifestations encountered in ordinary management. Thirdly, no additional testing was undertaken; hence, the model's predicted application to independently verified samples as well as distributed observations continues undocumented [21,22].

Further investigation should evaluate AI efficiency across information containing questionable OLP disorders, inflammation responses, and possibly abnormal diseases. The incorporation of multidisciplinary findings as well as outside testing will aid in determining the sustained performance and therapeutic application of artificial neural network models in a variety of pathogenic clinical situations. Furthermore, integrating AI into pathologists' evaluations might enhance workflow effectiveness while retaining excellent diagnosing confidence [23, 24].

5 CONCLUSION

Subsequently, our work demonstrates the fact that artificial intelligence (AI) might achieve 100% concurrence with professional oral examiners in identifying standardized OLPs. Despite the results above encouraging the use of computer vision as an additional diagnosing technique, additional evidence using diverse and multidisciplinary collections would be necessary to evaluate its reliability in standard clinical settings.

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Acknowledgments

We acknowledge the Ministry of health, Medical city and College of Dentistry of Baghdad University that help in manuscript preparation.

Disclosure of conflict of interest

I declare that there is no possible conflict of interest in the publication of this work. Furthermore, I have personally observed ethical difficulties like plagiarism, consent forms, misconduct, data fabrication and/or deception, multiple publishing and/or submission, and duplication.

Statement of informed consent

Informed consent was obtained from all individual participants included in the study.

Author contributions

All authors contributed in collected, analyzed, and interpreted the data, but also wrote, analyzed, and put together the paper.

REFERENCES

[1] Eisen D, Carrozzo M, Bagan Sebastian JV, Thongprasom K. Oral lichen planus: clinical features and management. Oral Dis. 2005;11(6):338–349.

[2] Alrashdan MS, Cirillo N, McCullough M. Oral lichen planus: a comprehensive review. Arch Dermatol Res. 2016;308(8):539–551.

[3] van der Meij EH, van der Waal I. Lack of clinicopathologic correlation in oral lichen planus. Oral Surg Oral Med Oral Pathol Oral Radiol Endod. 2003;95(4):447–452.

[4] Cheng YS, Gould A, Kurago Z, Fantasia J, Muller S. Diagnosis of oral lichen planus: a position paper. Oral Surg Oral Med Oral Pathol Oral Radiol. 2016;122(3):332–354.

[5] Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60–88.

[6] Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542:115–118.

[7] Aubreville M, Bertram CA, Marzahl C, et al. Deep learning in histopathology: a review. J Pathol Inform. 2020;11:20.

[8] Shaban MT, Baur C, Navab N, Albarqouni S. Context-aware AI for digital pathology. IEEE Trans Med Imaging. 2019;39(7):2394–2405.

[9] Madabhushi A, Lee G. Image analysis and machine learning in digital pathology. Med Image Anal. 2016;33:170–175.

[10] Komura D, Ishikawa S. Machine learning methods for histopathological image analysis. Comput Struct Biotechnol J. 2018;16:34–42.

[11] Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis. J Pathol Inform. 2016;7:29.

[12] Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology. Nat Rev Clin Oncol. 2019;16(11):703–715.

[13] Steiner DF, MacDonald R, Liu Y, et al. Impact of deep learning assistance on pathologist performance. NPJ Digit Med. 2018;1:23.

[14] Bejnordi BE, Veta M, van Diest PJ, et al. Diagnostic assessment of deep learning algorithms for detection of metastases. JAMA. 2017;318(22):2199–2210.

[15] Abels E, Pantanowitz L. Current state of the regulatory landscape for AI in pathology. J Pathol Inform. 2017;8:1–7.

[16] Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer. 2018;18(8):500–510.

[17] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44–56.

[18] Madabhushi A, Lee G. Image analysis and machine learning in digital pathology. Med Image Anal. 2016;33:170–175.

[19] Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis. J Pathol Inform. 2016;7:29.

[20] Komura D, Ishikawa S. Machine learning methods for histopathological image analysis. Comput Struct Biotechnol J. 2018;16:34–42.

[21] Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology. Nat Rev Clin Oncol. 2019;16(11):703–715.

[22] Steiner DF, MacDonald R, Liu Y, et al. Impact of deep learning assistance on pathologist performance. NPJ Digit Med. 2018;1:23.

[23] Bejnordi BE, Veta M, van Diest PJ, et al. Diagnostic assessment of deep learning algorithms for detection of metastases. JAMA. 2017;318(22):2199–2210.

[24] Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer. 2018;18(8):500–510.