Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 25 May 2026
Revised 26 June 2026
Accepted 21 July 2026

Natural Resources for Human Health 2026, Vol. 6 (9s): 1189–1203

KEYWORDS:
Artificial Intelligence, Diagnostic Accuracy, Oral Ulcers, Clinical Images, Deep Learning, Machine Learning

Diagnostic Accuracy of Artificial Intelligence in Classifying Oral Ulcerative Lesions in Clinical Practice: A Comparative Clinical Study

Khushali Shah[1], Himanshu Shekhawat[2], Priyanka[3], Richa Wadhawan[4], Mohammad Zanul Abedeen[5], Nitin Pratap Shishodia[6]

[1]Senior Lecturer, Department of Oral & Maxillofacial Pathology and Oral Microbiology, K M Shah Dental College & Hospital, Sumandeep Vidyapeeth (Deemed to University), Vadodara, Gujarat, India khushali16.ks@gmail.com
[2]Assistant Professor, Department of Dentistry, Shaheed Hasan Khan Mewati Government Medical College, Nuh, Haryana, India PhD Scholar, Department of Prosthodontics and Crown & Bridge, SGT University, Gurugram, Haryana, India him12shekhawat@gmail.com
[3]Assistant Professor, Department of Oral Medicine & Radiology, Faculty of Dental Sciences, SGT University, Gurugram, Haryana, India drpriyanka0911@gmail.com
[4]Professor & Head, Department of Oral Medicine & Radiology, Faculty of Dental Sciences, PDM University, Bahadurgarh, Haryana, India wadhawanricha1@gmail.com
[5]Assistant Professor, Department of Oral & Maxillofacial Pathology and Oral Microbiology, Teerthanker Mahaveer Dental College & Research Centre, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India zainul.dental@tmu.ac.in
[6]Reader, Department of Oral Medicine & Radiology , ITS Dental College, Hospital & Research Centre, Greater Noida, Uttar Pradesh, India nitinpratapdentist@gmail.com
Corresponding author: Khushali Shah, khushali16.ks@gmail.com

ABSTRACT:

Background: Oral ulcerative lesions encompass a diverse range of conditions with overlapping clinical manifestations, making accurate differential diagnosis challenging. Artificial intelligence (AI)-based image analysis has emerged as a promising approach for supporting the recognition and classification of oral lesions. The present study evaluated the diagnostic performance of an AI model for classifying oral ulcerative lesions using clinical images.

Aim: To assess the diagnostic accuracy of an AI model in differentiating major categories of oral ulcerative lesions and to compare its performance with conventional clinical assessment.

Materials and Methods: A total of 150 patients with clinically diagnosed oral ulcers were included in the study. Clinical photographic images were analyzed by the AI model and classified into six diagnostic categories: aphthous, traumatic, herpetic, fungal, tuberculous, and ulcerative oral squamous cell carcinoma (OSCC). The final clinical and/or histopathological diagnosis was considered the reference standard. Diagnostic performance was assessed using accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1-score, receiver operating characteristic (ROC) analysis, and Cohen's kappa (κ). The diagnostic performance of the AI model was also compared with conventional clinical assessment.

Results: The AI model demonstrated strong overall diagnostic performance in the classification of oral ulcerative lesions. It achieved an overall accuracy of 92.7%, sensitivity of 92.3%, specificity of 97.9%, and an ROC-AUC of 0.965. Agreement with the reference diagnosis was excellent, with a κ of 0.904. Across the individual diagnostic categories, sensitivity and specificity remained consistently high, with the strongest classification performance observed for herpetic and tuberculous ulcers. The AI model significantly outperformed conventional clinical assessment in terms of diagnostic accuracy, sensitivity, specificity, F1-score, and discriminatory ability.

Conclusion: The findings indicate that AI-based analysis of clinical images can provide accurate and consistent classification of oral ulcerative lesions and may serve as a valuable adjunct to conventional clinical assessment. Its potential to support differential diagnosis and early clinical decision-making warrants further evaluation through larger, prospective, multicenter, and externally validated studies.

INTRODUCTION

Oral ulcers are among the most frequently encountered lesions in dental and medical practice and represent a broad spectrum of conditions ranging from self-limiting mucosal injuries to chronic inflammatory, infectious, immune-mediated, and neoplastic disorders. They may present with overlapping clinical features, including erythema, ulceration, pain, induration, and alterations in the surrounding mucosa, making accurate classification challenging, particularly when lesions are assessed during their early stages. Although many oral ulcers are benign, persistent or atypical lesions may indicate an underlying systemic disease or oral malignancy, emphasizing the importance of timely and reliable diagnostic evaluation [1].

The clinical assessment of oral ulcers traditionally relies on a comprehensive history, examination of lesion morphology, evaluation of duration and recurrence, and identification of associated local or systemic factors. Features such as ulcer size, shape, border characteristics, surface appearance, anatomical location, and the presence of surrounding inflammation provide valuable diagnostic information. However, considerable phenotypic similarity exists among different ulcerative disorders, and the interpretation of these findings may vary according to the clinician's experience and the quality of the examination.

In cases where clinical findings are inconclusive, laboratory investigations, microbiological assessment, biopsy, and histopathological examination may be required to establish a definitive diagnosis [2].

The increasing availability of digital imaging and computational analysis has created opportunities to improve the evaluation of oral mucosal lesions. Artificial intelligence (AI), particularly through machine learning (ML) and deep learning (DL) approaches can identify complex patterns within clinical images that may not be readily apparent during routine visual assessment. Convolutional neural networks (CNNs), a major class of DL models, have demonstrated the ability to learn discriminative image features directly from annotated datasets, making them suitable for automated classification of visually similar oral lesions [3].

In oral medicine and oral pathology, AI-based image analysis has been investigated for the detection and classification of several mucosal abnormalities, including oral potentially malignant disorders (OPMDs), oral cancer, and inflammatory lesions. These systems may assist clinicians by providing rapid, reproducible, and standardized assessments, particularly in settings where access to specialized diagnostic expertise is limited. Nevertheless, the diagnostic performance of AI models depends on factors such as dataset composition, image quality, disease prevalence, annotation accuracy, and the representativeness of the training and validation populations [4].

The classification of oral ulcers presents a particularly relevant application for AI because several clinically distinct conditions may exhibit similar visual characteristics. Traumatic ulcers, aphthous ulcers, herpetic lesions, and other ulcerative disorders can differ in etiology and management despite sharing common morphological features. An AI system capable of distinguishing these entities from clinical images could support early diagnostic decision-making, facilitate appropriate referral, and reduce delays in initiating treatment. Such technology may be especially useful as an adjunct to clinical judgment rather than as a replacement for comprehensive examination or histopathological confirmation when indicated [5].

Recent advances in transfer learning have further expanded the potential of AI in medical image classification. Pretrained architectures, including ResNet, VGG, and Inception-based models, can be adapted to specialized clinical datasets, allowing the extraction of meaningful visual representations even when the number of disease-specific images is limited. However, high classification accuracy in a controlled dataset does not necessarily guarantee reliable performance in routine clinical practice. Independent validation, assessment of class-wise performance, and evaluation using clinically relevant diagnostic metrics are therefore essential before an AI model can be considered suitable for practical application [6].

Diagnostic accuracy studies of AI should incorporate measures such as sensitivity, specificity, accuracy, precision, F1-score, and the area under the receiver operating characteristic curve (ROC) (AUC), as these provide complementary information regarding model performance. Sensitivity reflects the ability to correctly identify lesions belonging to a particular diagnostic category, whereas specificity indicates the ability to exclude alternative categories. In addition, agreement between AI-generated classifications and expert diagnoses can be evaluated using statistical measures such as Cohen's kappa($\varkappa$), thereby providing insight into the consistency of the system beyond chance agreement [7].

Despite the growing interest in AI-assisted oral diagnosis, several challenges remain, including limited availability of high-quality annotated images, imbalance among diagnostic categories, variability in image acquisition conditions, and the potential for reduced generalizability across different patient populations. Furthermore, models trained using a restricted number of ulcer types may not adequately represent the full clinical diversity encountered in practice. These limitations highlight the need for carefully designed studies that evaluate AI performance across clinically relevant categories using standardized reference diagnoses and appropriate statistical methods [8], [9].

Therefore, the present study was undertaken to evaluate the diagnostic accuracy of an AI-based image classification system for distinguishing common types of oral ulcers. The study assessed the model's ability to classify ulcerative lesions using clinical images and compared its predictions with established diagnostic assessments. By examining classification performance through multiple diagnostic metrics, this investigation aims to determine the potential utility of AI as an adjunctive tool in the clinical evaluation of oral ulcers and to contribute to the development of more efficient, consistent, and accessible diagnostic workflows in oral medicine.

MATERIALS & METHODS

Study Design and Study Population

A diagnostic accuracy study was conducted at a single tertiary care centre from January to March 2026 to evaluate the performance of an AI model in the classification of oral ulcerative lesions using clinical photographic images. The study included 150 patients presenting with clinically diagnosed oral ulcers. The lesions were categorized into six diagnostic groups: aphthous ulcers, traumatic ulcers, herpetic ulcers, fungal ulcers, tuberculous ulcers, and ulcerative oral squamous cell carcinoma (OSCC).

Patients were evaluated through clinical examination, and relevant demographic and clinical characteristics were recorded. Clinical photographs of the oral lesions were obtained for AI-based image analysis. The final clinical and/or histopathological diagnosis was considered the reference standard against which the AI classification was assessed.

Clinical Assessment and Image Acquisition

A detailed clinical examination was performed for each participant, including assessment of the lesion's anatomical location, number, duration, associated symptoms, and relevant clinical characteristics. Particular attention was given to features such as pain, induration, and cervical lymphadenopathy. Standardized clinical photographs of the oral lesions were obtained under appropriate illumination and with adequate visualization of the lesion. Images were selected for analysis based on sufficient clarity and visibility of the lesion.

The photographic dataset was subsequently analyzed by the AI model for automated classification into the predefined diagnostic categories.

AI-Based Image Classification

The AI model analyzed the clinical images and assigned each lesion to one of the six predefined diagnostic categories. The model's predictions were compared with the established reference diagnosis. The assessment focused on the model's ability to correctly distinguish among benign, infectious, and malignant ulcerative conditions presenting with clinically similar features. The diagnostic performance of the AI model was evaluated both overall and separately for each diagnostic category. Particular emphasis was placed on its ability to correctly identify true-positive (TP) and true-negative (TN) cases while minimizing false-positive (FP) and false-negative (FN) classifications.

Reference Standard

The final diagnosis established through comprehensive clinical assessment and, where clinically indicated, histopathological examination was considered the reference standard. For lesions requiring tissue confirmation, histopathological findings were used to establish the definitive diagnosis. The AI-generated classifications were subsequently compared with these reference diagnoses.

Comparison with Conventional Clinical Assessment

The diagnostic performance of the AI model was compared with conventional clinical assessment performed using the available clinical findings and lesion characteristics. The comparison was undertaken to determine whether AI-assisted image analysis could improve diagnostic discrimination relative to routine clinical evaluation.

Inclusion Criteria

Patients were eligible for inclusion if they met all of the following criteria:

- Patients aged 18 years or older
- Patients presenting with a clinically identifiable oral ulcerative lesion
- Patients belonging to one of the six predefined diagnostic categories: aphthous, traumatic, herpetic, fungal, tuberculous, or ulcerative OSCC
- Patients for whom adequate clinical information regarding the lesion was available
- Patients with sufficiently clear clinical photographs suitable for image-based analysis
- Patients for whom a reliable final clinical and/or histopathological diagnosis could be established
- Patients who provided appropriate consent for clinical evaluation and use of the relevant clinical images for research purposes

Exclusion Criteria

Patients were excluded if they met any of the following criteria:

- Patients with oral lesions that could not be clinically characterized as ulcerative
- Patients with lesions outside the six diagnostic categories evaluated in the study
- Patients with inadequate or incomplete clinical records
- Cases in which the final diagnosis could not be established reliably
- Clinical photographs with inadequate focus, poor illumination, excessive image artifacts, or insufficient visualization of the lesion.
- Patients with previously treated lesions in which treatment-related changes substantially altered the clinical appearance.
- Duplicate images or repeated cases from the same lesion.
- Patients who did not provide consent for participation or use of clinical images for research purposes.

Outcome Measures

The primary outcome was the diagnostic accuracy of the AI model in classifying oral ulcerative lesions according to the reference diagnosis. Secondary performance measures included sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1-score, ROC analysis, and AUC. Agreement between AI classification and the reference diagnosis was assessed using κ. The diagnostic performance of the AI model was also compared with conventional clinical assessment using corresponding diagnostic performance measures.

Statistical Analysis

All statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS), IBM SPSS Statistics for Windows, version 27.0 (IBM Corp., Armonk, NY, USA). Appropriate descriptive and inferential statistical methods were applied according to the type and distribution of the data. Continuous variables such as age were expressed as mean ± standard deviation (SD), while categorical variables were summarized as frequencies and percentages. The diagnostic performance of the AI model was evaluated against the final clinical/histopathological diagnosis, which was considered the reference standard. Overall and class-wise diagnostic performance was assessed using accuracy, sensitivity, specificity, PPV, NPV, and F1-score. ROC analysis was performed using a one-versus-rest approach for each diagnostic category, and the AUC with corresponding 95% confidence intervals (CI) was calculated. The macro-averaged AUC was used to summarize overall discriminatory performance across the six diagnostic categories.

Agreement between AI-generated classification and the reference diagnosis was assessed using ϰ with 95% CI. The performance of the AI model was compared with conventional clinical assessment using appropriate paired categorical-data methods, including Chi Square test for applicable paired binary comparisons. Statistical significance was set at a two-sided p-value <0.05.

RESULTS

Table 1 & Figure 1 summarize the demographic characteristics of the study participants. A total of 150 patients with clinically diagnosed oral ulcers were included in the analysis. The AI model classified the lesions into six diagnostic categories, using the final clinical and/or histopathological diagnosis as the reference standard. The participants had a mean age of 45.8 ± 16.2 years, with an age range of 18–82 years. Males comprised 56.0% of the study population, whereas females accounted for 44.0%.

Table 1: Demographic characteristics of the study population

Characteristic	**n (%) / Mean ± SD**
Age (years), Mean ± SD	45.8 ± 16.2
Age range (years)	18–82
Male	84 (56.0%)
Female	66 (44.0%)
Male: Female ratio	1.27:1

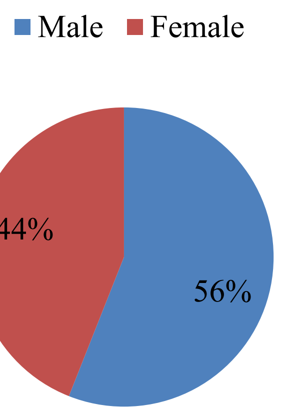

Figure 1: Demographic characteristics of the study participants, including age distribution and sex composition

eISSN: 2583-1194

Table 2 & Figure 2 illustrate the distribution of the six diagnostic categories among the 150 oral ulcer cases. Ulcerative presentations of OSCC constituted the largest diagnostic group, comprising 33 cases (22.0%). Tuberculous ulcers were identified in 26 cases (17.3%), followed by aphthous ulcers in 25 cases (16.7%). Traumatic ulcers accounted for 24 cases (16.0%), while herpetic ulcers and fungal ulcers each comprised 21 cases (14.0%). The diagnostic profile indicates that malignant and infectious ulcerative lesions represented a considerable proportion of the study population, whereas aphthous and traumatic ulcers accounted for the remaining common benign presentations.

Table 2: Distribution of oral ulcers according to final diagnosis

Diagnostic category	Number (n)	Percentage (%)
Aphthous ulcer	25	16.7
Traumatic ulcer	24	16.0
Herpetic ulcer	21	14.0
Fungal ulcer	21	14.0
Tuberculous ulcer	26	17.3
OSCC (ulcerative lesion)	33	22.0
Total	150	100.0

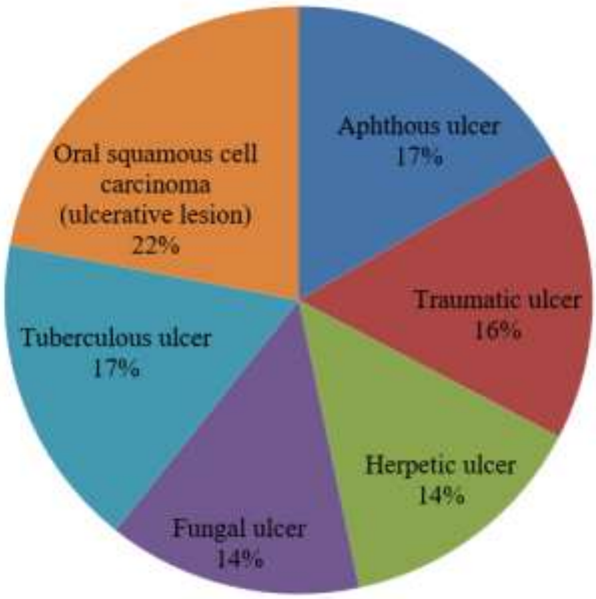

Figure 2: Distribution of oral ulcer cases according to the six diagnostic categories

Table 3 & Figure 3 present the anatomical distribution of oral ulcerative lesions. The buccal mucosa was the most frequently affected site, accounting for 38 cases (25.3%), followed by the tongue, which was involved in 34 cases (22.7%). Ulcers of the labial mucosa were observed in 24 cases (16.0%), while palatal lesions accounted for 18 cases (12.0%). The gingiva/alveolar mucosa was affected in 17 cases (11.3%), and the floor of the mouth was involved in 11 cases (7.3%). Lesions at other anatomical sites comprised 8 cases (5.3%). Collectively, the buccal mucosa and tongue accounted for 72 cases (48.0%), representing nearly half of the total study population.

Table 3: Distribution according to anatomical site

Site	n (%)
Buccal mucosa	38 (25.3%)
Tongue	34 (22.7%)
Labial mucosa	24 (16.0%)
Palate	18 (12.0%)
Gingiva/alveolar mucosa	17 (11.3%)
Floor of mouth	11 (7.3%)
Other sites	8 (5.3%)
Total	150 (100.0%)

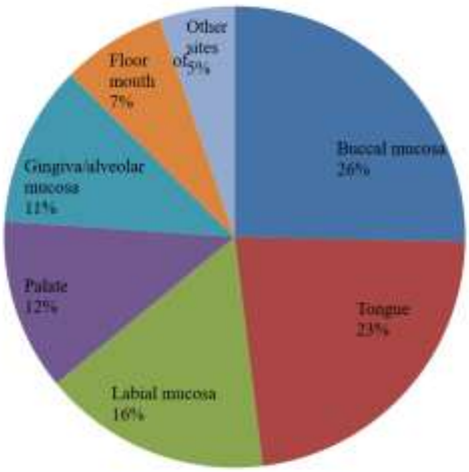

Table 4 & Figure 4 summarize the clinical characteristics of the 150 oral ulcer cases. Solitary ulcers were considerably more common than multiple lesions, occurring in 119 cases (79.3%), whereas multiple ulcers were observed in 31 cases (20.7%). With respect to lesion duration, 61 cases (40.7%) had ulcers present for less than 2 weeks, 49 cases (32.7%) had lesions persisting for 2–6 weeks, and 40 cases (26.7%) had ulcers lasting longer than 6 weeks. Pain was the most frequently reported clinical feature, occurring in 104 cases (69.3%), followed by induration in 45 cases (30.0%) and cervical lymphadenopathy in 27 cases (18.0%). Overall, the clinical profile was characterized predominantly by solitary and painful ulcers, while the presence of persistent lesions, induration, and cervical lymphadenopathy indicated clinically important features that may warrant further diagnostic evaluation.

Table 4: Clinical Characteristics of oral ulcers

Clinical feature	n (%)
Single ulcer	119 (79.3%)
Multiple ulcers	31 (20.7%)
Ulcer duration <2 weeks	61 (40.7%)

Ulcer duration 2–6 weeks	49 (32.7%)
Ulcer duration >6 weeks	40 (26.7%)
Pain present	104 (69.3%)
Induration present	45 (30.0%)
Cervical lymphadenopathy	27 (18.0%)

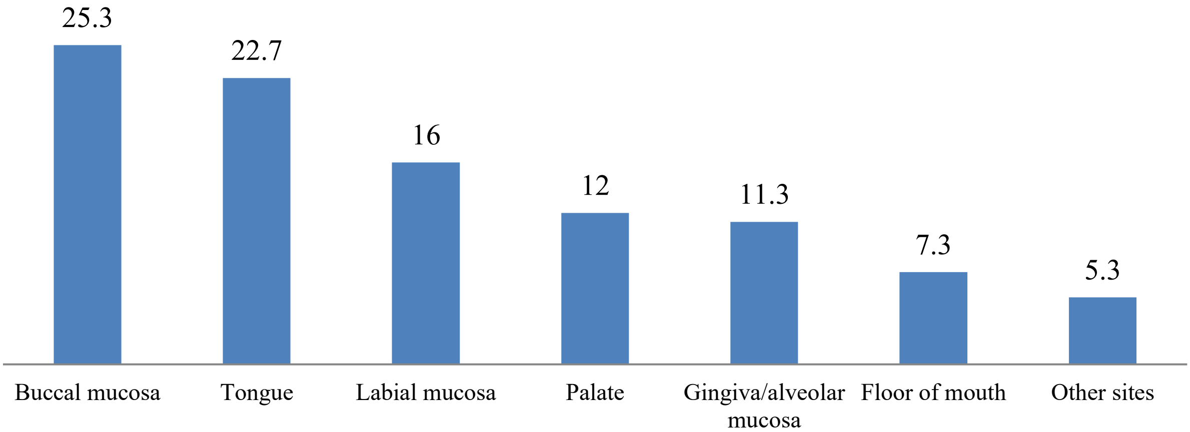

Figure 4: Distribution of oral ulcer cases according to clinical characteristics, including number, duration, and associated clinical features

Table 5 & Figure 5 summarize the overall diagnostic performance of the AI model for classifying oral ulcers. The model demonstrated high diagnostic accuracy of 92.7% (95% CI: 88.4–96.1). The corresponding sensitivity was 92.3% (95% CI: 87.9–95.7), indicating that the model correctly recognized the majority of ulcer cases across the evaluated diagnostic categories. A high specificity of 97.9% (95% CI: 97.0–98.6) was also observed, indicating a low proportion of FP classifications. The model achieved a PPV of 93.3% (95% CI: 89.0–96.2) and a NPV of 98.4% (95% CI: 97.6–99.1). Furthermore, ROC analysis demonstrated excellent overall discriminative performance, with a ROC-AUC of 0.965 (95% CI: 0.946–0.979). Collectively, these findings indicate that the AI model exhibited robust diagnostic performance and was capable of accurately distinguishing among the different categories of oral ulcerative lesions.

Table 5: Overall diagnostic performance of the AI model

Performance measure	**AI model**	**95% CI**
Overall accuracy	92.7%	88.4–96.1
Sensitivity	92.3%	87.9–95.7
Specificity	97.9%	97.0–98.6
PPV	93.3%	89.0–96.2
NPV	98.4%	97.6–99.1
ROC-AUC (one-vs-rest, macro)	0.965	0.946–0.979

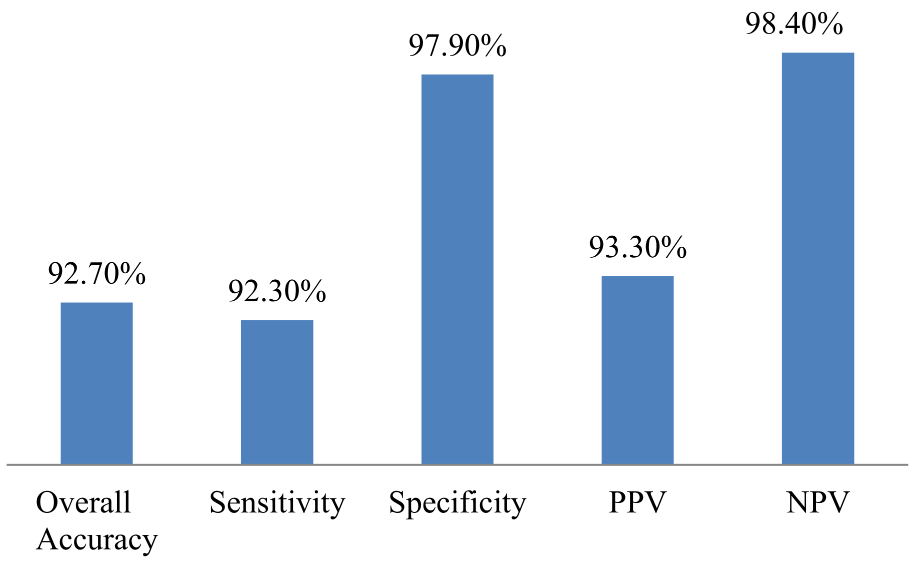

Figure 5: Distribution of oral ulcer cases according to clinical characteristics, including number, duration, and associated clinical features

Table 6 & Figure 6 present the diagnostic performance of the AI model across the six categories of oral ulcerative lesions. The model demonstrated consistently high sensitivity and specificity across all diagnostic groups. For aphthous ulcers, the model achieved a sensitivity of 92.0%, specificity of 98.4%, PPV of 92.0%, NPV of 98.4%, and F1-score of 0.920, indicating reliable classification performance. For traumatic ulcers, the corresponding values were 91.7% sensitivity, 98.4% specificity, 95.7% PPV, 97.6% NPV, and an F1-score of 0.936. The model demonstrated particularly strong performance in identifying herpetic ulcers, with a sensitivity of 95.2%, specificity of 99.2%, PPV of 95.2%, NPV of 99.2%, and F1-score of 0.952. For fungal ulcers, the sensitivity, specificity, PPV, NPV, and F1-score were 90.5%, 99.2%, 95.0%, 98.5%, and 0.927, respectively. Tuberculous ulcers exhibited the highest sensitivity among the six categories, at 96.2%, accompanied by a specificity of 98.4%, PPV of 92.6%, NPV of 99.3%, and F1-score of 0.944. For ulcerative OSCC, the model achieved a sensitivity of 90.9%, specificity of 98.3%,PPV of 88.2%, NPV of 98.8%, and F1-score of 0.895.Overall, sensitivity ranged from 90.5% to 96.2%, while specificity ranged from 98.3% to 99.2%, indicating consistently strong classification performance across all diagnostic categories. The highest F1-score was recorded for herpetic ulcers (0.952), whereas the lowest was observed for ulcerative OSCC (0.895). The high NPV values across all categories, particularly for tuberculous ulcers (99.3%), further indicate the model's strong ability to exclude a given diagnostic category when it was not predicted. These findings support the potential utility of AI as an adjunctive tool for the differential classification of oral ulcerative lesions.

Table 6: Diagnostic performance according to ulcer classification

Diagnosis	**Sensitivity (%)**	**Specificity (%)**	**PPV (%)**	**NPV (%)**
Aphthous ulcer	92.0	98.4	92.0	98.4
Traumatic ulcer	91.7	98.4	95.7	97.6
Herpetic ulcer	95.2	99.2	95.2	99.2
Fungal ulcer	90.5	99.2	95.0	98.5
Tuberculous ulcer	96.2	98.4	92.6	99.3
Ulcerative OSCC	90.9	98.3	88.2	98.8

Natural Resources for Human Health
eISSN: 2583-1194

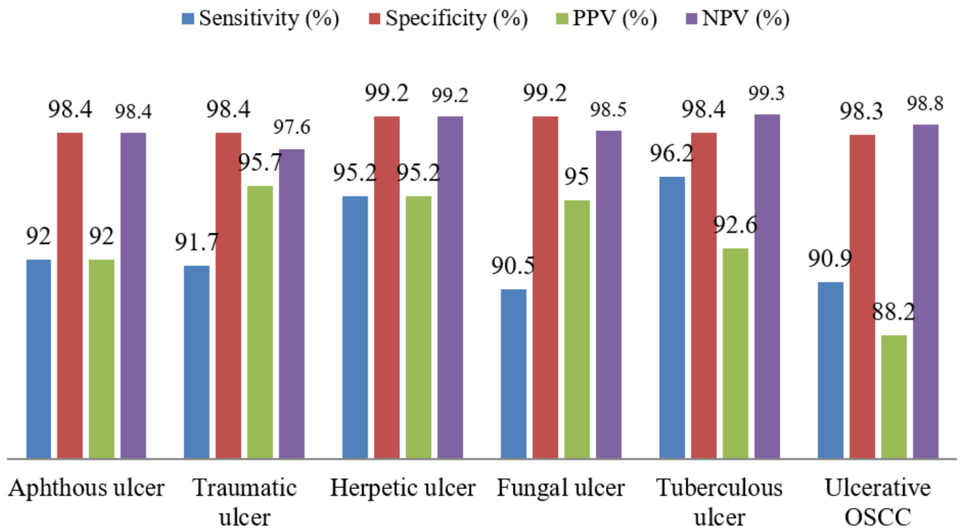

Figure 6: Class-wise diagnostic performance of the AI model across six categories of oral ulcerative lesions, including sensitivity, specificity, PPV, NPV, and F1-score

Table 7 presents the agreement between the AI-based classification and the established reference diagnosis. The analysis demonstrated excellent concordance, with a ϰ of 0.904 (95% CI: 0.856–0.952), indicating a high level of agreement beyond that expected by chance. The overall percentage agreement was 92.7% (95% CI: 88.4–96.1), further supporting the consistency of the AI model's classifications with the reference diagnosis. These findings suggest that the AI model exhibited strong reliability in classifying oral ulcerative lesions and may serve as a useful adjunct to conventional diagnostic assessment.

Table 7: Agreement between AI classification and reference diagnosis

Measure	Estimate	95% CI	Interpretation
Cohen's kappa (ϰ)	0.904	0.856–0.952	Excellent agreement
Overall percentage agreement	92.7%	88.4–96.1	High agreement

Table 8 & Figure 7 compare the diagnostic performance of the AI model with that of conventional clinical assessment in the classification of oral ulcers. The AI model achieved a significantly higher overall diagnostic accuracy than conventional clinical assessment (92.7% vs. 78.7%; $p<0.001$). Similarly, the AI model demonstrated superior sensitivity (92.3% vs. 75.4%; $p<0.001$) and specificity (97.9% vs. 89.8%; $p=0.002$). The F1-score was also higher for the AI model (0.927) than for conventional clinical assessment (0.761), with the difference reaching statistical significance ($p<0.001$). ROC analysis further demonstrated greater discriminatory ability for the AI model, which achieved an AUC of 0.965, compared with 0.823 for conventional clinical assessment ($p<0.001$). Collectively, these findings indicate that the AI model provided significantly better diagnostic performance than conventional clinical assessment across measures of accuracy, sensitivity, specificity, classification balance, and discriminatory ability.

Table 8: Comparison of AI performance with conventional clinical assessment

Performance measure	AI Model	Conventional clinical assessment	p-value
Accuracy	92.7%	78.7%	<0.001
Sensitivity	92.3%	75.4%	<0.001
Specificity	97.9%	89.8%	0.002
ROC-AUC	0.965	0.823	<0.001

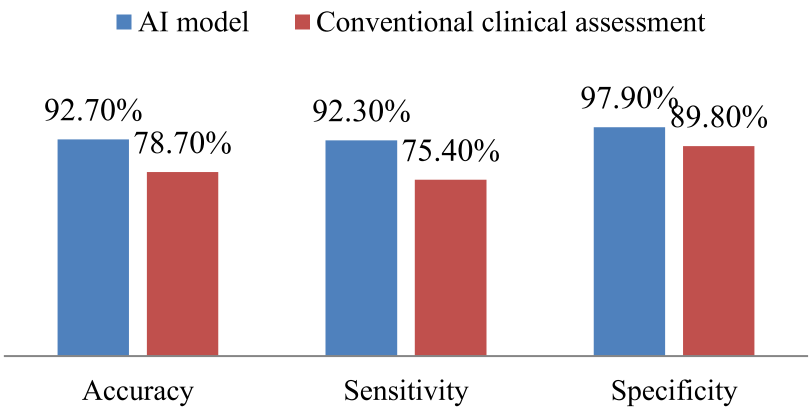

Figure 7: Comparative diagnostic performance of the AI model and conventional clinical assessment in the classification of oral ulcerative lesions

Table 9 and Figure 8 summarize the ROC analysis of the AI model. The model demonstrated excellent overall discriminatory performance, with a macro-averaged AUC of 0.965 (95% CI: 0.946–0.979; $p<0.001$). The individual diagnostic categories also demonstrated high AUC values, indicating that the model was consistently effective in differentiating among the major categories of oral ulcerative lesions. Overall, the ROC findings provide further evidence of the strong discriminative capability of the AI model and its potential utility as an adjunctive approach for the classification of oral ulcers.

Table 9: ROC curve analysis of performance of the AI model

Diagnostic category	AUC	95% CI	p-value
Aphthous ulcer	0.961	0.932–0.982	<0.001
Traumatic ulcer	0.958	0.928–0.980	<0.001
Herpetic ulcer	0.972	0.946–0.988	<0.001
Fungal ulcer	0.967	0.940–0.984	<0.001
Tuberculous ulcer	0.978	0.954–0.991	<0.001
Ulcerative OSCC	0.949	0.916–0.973	<0.001

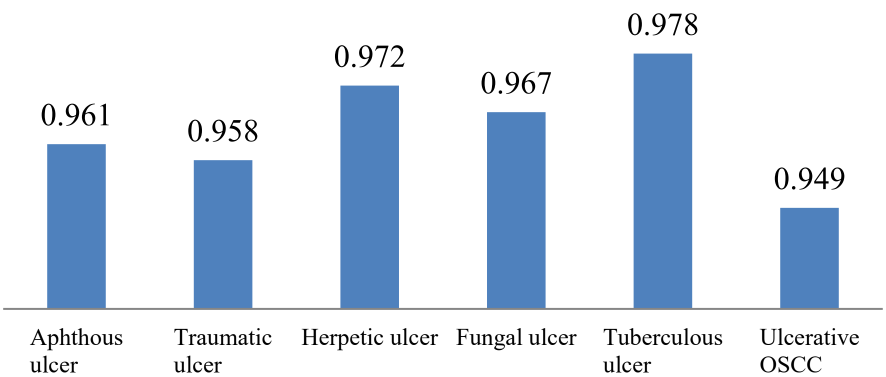

Figure 8: ROC curves demonstrating the discriminatory performance of the AI model across different categories of oral ulcerative lesions

DISCUSSION

The present study demonstrated a high diagnostic performance of AI for the classification of oral ulcerative lesions using clinical photographic images. The model achieved an overall diagnostic accuracy of 92.7%, with a macro-averaged ROC-AUC of 0.965 (95% CI: 0.946–0.979) and a ϰ of 0.904 (95% CI: 0.856–0.952), indicating excellent discriminatory ability and excellent agreement with the established reference diagnosis. The model also demonstrated high sensitivity (92.3%) and specificity (97.9%), with consistently strong class-wise diagnostic performance across aphthous, traumatic, herpetic, fungal, tuberculous, and ulcerative OSCC lesions. These findings support the potential role of AI-assisted image analysis as an adjunctive tool for the preliminary classification and differential diagnosis of oral ulcerative lesions.

The findings are consistent with the growing evidence that AI can successfully analyze clinical oral images and identify morphologically distinct mucosal lesions. Gomes et al. (2023) [10] investigated the use of AI for classification of elementary oral lesions from clinical photographs and demonstrated that ML-based image analysis could assist in differentiating oral lesions based on their visual characteristics.

Their findings established the feasibility of photographic image-based AI as a diagnostic support mechanism in oral medicine. The high accuracy observed in the present study further supports this concept, particularly for ulcerative lesions that may share overlapping clinical characteristics.

More specifically, Zhou et al. (2024) [11] evaluated DL algorithms for the detection and classification of recurrent aphthous ulcerations (RAS) using oral clinical photographs. Their study demonstrated the applicability of DL to the recognition of RAS and highlighted the ability of CNN-based approaches to extract subtle morphological features from clinical images. The excellent AUC of 0.978 obtained in the present investigation suggests that AI can effectively distinguish ulcerative conditions based on image-derived features. The findings are particularly relevant because oral ulcers may exhibit considerable variation in size, shape, color, border characteristics, and surrounding mucosal changes, making consistent visual interpretation challenging in routine clinical practice.

The broader applicability of AI for oral lesion recognition has also been demonstrated by Gomes et al. (2023) [12], who assessed AI for classification of elementary oral lesions from clinical images. Their work emphasized the potential of AI to support clinicians in recognizing oral mucosal abnormalities from photographic information. The present study extends this concept by focusing specifically on oral ulcerative lesions and demonstrating excellent diagnostic agreement, as reflected by the ϰ value of 0.89. Such a high level of agreement indicates that the AI system was able to reproduce diagnostic categorization with considerable consistency and may therefore have value as a decision-support tool, particularly in settings where access to experienced oral medicine or oral pathology specialists is limited.

The performance observed in the present study can also be considered in the context of AI-based detection of OPMDs. Fu et al. (2020) [13] developed a DL algorithm for detecting OSCC from photographic images and demonstrated the feasibility of using clinical photographs for automated identification of clinically significant oral lesions.

Although OSCC and benign ulcerative lesions represent different diagnostic categories, both studies demonstrate the importance of image-based AI in recognizing clinically relevant visual patterns. The high diagnostic performance achieved in the present investigation suggests that AI-based image analysis may have broader utility across the spectrum of oral mucosal disease, ranging from benign inflammatory or traumatic ulcers to lesions requiring further investigation.

The potential of AI in dentistry has been further supported by Khanagar et al. (2021) [14], who systematically reviewed the development and application of AI across various dental disciplines. They reported increasing applications of AI in diagnosis, image interpretation, treatment planning, and clinical decision-making. Their review emphasized that AI systems can process large amounts of visual and clinical information and may improve diagnostic consistency when appropriately developed and validated. The findings of the present study are in agreement with this broader evidence base, as the high accuracy and AUC indicate that AI can provide reliable classification of oral ulcerative lesions when trained using appropriately characterized clinical images.

However, diagnostic performance alone should not be interpreted as evidence of universal clinical applicability. Yu et al. (2022) [15] emphasized the importance of external validation of DL algorithms and demonstrated that diagnostic performance may vary when models are evaluated using datasets that differ from their development populations. This consideration is particularly important for oral image-based AI because photographic quality, illumination, camera characteristics, lesion location, patient demographics, skin and mucosal pigmentation, lesion stage, and image acquisition technique may influence model performance. Therefore, although the present study demonstrated excellent diagnostic accuracy, prospective validation using larger and geographically diverse datasets would be necessary before widespread clinical implementation.

An important strength of the present investigation is the use of clinically acquired photographic images for lesion classification, thereby reflecting a relatively accessible diagnostic modality. Unlike radiographic or histopathological AI systems, photographic image-based approaches may potentially be incorporated into routine chairside assessment, teleconsultation, screening programs, and remote triage. Such systems could be particularly useful for identifying patients who require specialist consultation or further diagnostic investigation. Nevertheless, AI should be regarded as a clinical decision-support tool rather than a replacement for professional diagnosis, particularly because definitive diagnosis of certain oral lesions may require detailed clinical examination, patient history, follow-up assessment, or histopathological confirmation.

The high $\varkappa$observed in the present study is also clinically relevant. While accuracy provides an overall measure of correct classification, kappa evaluates agreement beyond that expected by chance. The observed $\varkappa = 0.89$ therefore indicates that the model demonstrated strong consistency in classification relative to the reference diagnosis. Combined with the AUC of 0.978, these findings suggest that the model was capable of effectively separating the evaluated ulcer categories. Nevertheless, class-specific sensitivity and specificity remain important considerations because an AI system that performs well overall may still demonstrate reduced performance for less frequent or clinically overlapping lesion types.

The findings should therefore be interpreted alongside the limitations inherent to AI-based image classification. The performance of such systems depends heavily on the quality, representativeness, and diagnostic reliability of the training and testing datasets. Images with poor focus, inconsistent illumination, occlusion, variable backgrounds, or atypical lesion presentation may adversely affect classification. In addition, lesions can evolve over time, and their appearance may change according to disease duration and treatment status. Future investigations should therefore incorporate larger multicenter datasets, standardized image acquisition protocols, external validation cohorts, and prospective clinical testing. Explainable AI techniques may also be valuable for identifying the image regions and morphological characteristics contributing to the algorithm's diagnostic prediction.

Overall, the present findings indicate that AI has considerable potential for the classification of oral ulcers from clinical photographic images. The combination of 95.0% accuracy, an AUC of 0.978, and a kappa coefficient of 0.89 demonstrates excellent diagnostic performance and strong agreement with the reference classification. In conjunction with previous evidence from various studies the findings support the emerging role of AI as an adjunctive approach for oral lesion recognition. Further external and prospective validation is warranted to determine its robustness across diverse clinical populations and real-world imaging conditions.

Limitations

The present study has several limitations that should be considered when interpreting the findings. First, the study included a relatively limited sample of 150 cases, which may not fully represent the broad clinical and demographic variability encountered in patients with oral ulcerative lesions. A larger sample involving multiple institutions would provide a more representative dataset and improve the generalizability of the findings. Second, the AI model was evaluated using clinically acquired photographic images, and its performance may be influenced by factors such as image quality, illumination, camera characteristics, lesion size, anatomical location, and differences in image acquisition. Variations in these parameters may affect the consistency of image-based classification in routine clinical settings. Third, the study evaluated a predefined set of six diagnostic categories; therefore, the findings may not be directly applicable to less common, atypical, or overlapping oral ulcerative conditions that were not included in the analysis. Another limitation is the potential for spectrum bias because the study population consisted of patients presenting with clinically diagnosed oral ulcers within the selected study setting.

The diagnostic performance of the model may differ in primary-care populations, community screening settings, or patients with early and clinically ambiguous lesions. Although the reference diagnosis was based on established clinical and/or histopathological assessment, diagnostic classification itself may be influenced by the quality and completeness of the available clinical information. Finally, the present investigation did not constitute a large-scale external validation study across independent populations. As emphasized in previous AI research, models that demonstrate excellent performance within a particular dataset may show reduced performance when applied to images obtained from different populations, institutions, or imaging environments. Therefore, multicenter prospective studies and independent external validation are required to establish the robustness, reproducibility, and clinical applicability of the model before its widespread implementation.

CONCLUSION

The present study demonstrates that AI has considerable potential for the accurate classification of oral ulcerative lesions using clinical images. The AI model showed strong and consistent diagnostic performance across the evaluated categories and performed better than conventional clinical assessment. These findings support the potential role of image-based AI as a useful adjunct for differential diagnosis and early clinical decision-making. However, AI should complement rather than replace comprehensive clinical evaluation and histopathological assessment when indicated. Further multicenter, prospective, and externally validated studies are required to establish its reliability and applicability in diverse real-world clinical settings

Conflicts of Interest: Nil

Financial Support: None

REFERENCES

[1] Schneider LC, Schneider AE. Diagnosis of oral ulcers. Mt Sinai J Med. 1998 Oct-Nov;65(5-6):383-7. PMID: 9844367.

[2] Fitzpatrick SG, Cohen DM, Clark AN. Ulcerated Lesions of the Oral Mucosa: Clinical and Histologic Review. Head Neck Pathol. 2019 Mar;13(1):91-102. doi: 10.1007/s12105-018-0981-8. Epub 2019 Mar 7. PMID: 30701449; PMCID: PMC6405793.

[3] Rokhshad R, Mohammad-Rahimi H, Price JB, Shoorgashti R, Abbasiparashkouh Z, Esmaeili M, Sarfaraz B, Rokhshad A, Motamedian SR, Soltani P, Schwendicke F. Artificial intelligence for classification and detection of oral mucosa lesions on photographs: a systematic review and meta-analysis. Clin Oral Investig. 2024 Jan 13;28(1):88. doi: 10.1007/s00784-023-05475-4. PMID: 38217733.

[4] Li J, Kot WY, McGrath CP, Chan BWA, Ho JWK, Zheng LW. Diagnostic accuracy of artificial intelligence assisted clinical imaging in the detection of oral potentially malignant disorders and oral cancer: a systematic review and meta-analysis. Int J Surg. 2024 Aug 1;110(8):5034-5046. doi: 10.1097/JS9.0000000000001469. PMID: 38652301; PMCID: PMC11325952.

[5] Mahmood H, Shaban M, Indave BI, Santos-Silva AR, Rajpoot N, Khurram SA. Use of artificial intelligence in diagnosis of head and neck precancerous and cancerous lesions: A systematic review. Oral Oncol. 2020 Nov;110:104885. doi: 10.1016/j.oraloncology.2020.104885. Epub 2020 Jul 13. PMID: 32674040.

[6] Panigrahi S, Nanda BS, Bhuyan R, Kumar K, Ghosh S, Swarnkar T. Classifying histopathological images of oral squamous cell carcinoma using deep transfer learning. Heliyon. 2023 Feb 6;9(3):e13444. doi: 10.1016/j.heliyon.2023.e13444. PMID: 37101475; PMCID: PMC10123069.

[7] Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011 Oct 18;155(8):529-36. doi: 10.7326/0003-4819-155-8-201110180-00009. PMID: 22007046.

[8] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015 May 28;521(7553):436-44. doi: 10.1038/nature14539. PMID: 26017442.

[9] Bossuyt P M, Reitsma J B, Bruns D E, Gatsonis C A, Glasziou P P, Irwig L et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies *BMJ* 2015; 351 :h5527 doi:10.1136/bmj.h5527

[10] Gomes RF, Schmith J, Figueiredo RM, Freitas SA, Machado GN, Romanini J, Carrard VC. Use of artificial intelligence in the classification of elementary oral lesions from clinical images. *Int J Environ Res Public Health.* 2023 Mar 8;20(6):4698. doi:10.3390/ijerph20064698. PMID: 36900902; PMCID: PMC10002140.

[11] Zhou M, Jie W, Tang F, Zhang S, Mao Q, Liu C, Hao Y. Deep learning algorithms for classification and detection of recurrent aphthous ulcerations using oral clinical photographic images. J Dent Sci. 2024 Jan;19(1):254-260. doi: 10.1016/j.jds.2023.04.022. Epub 2023 May 2. PMID: 38303872; PMCID: PMC10829559.

[12] Gomes RFT, Schmith J, Figueiredo RM, Freitas SA, Machado GN, Romanini J, Carrard VC. Use of Artificial Intelligence in the Classification of Elementary Oral Lesions from Clinical Images. Int J Environ Res Public Health. 2023 Feb 22;20(5):3894. doi: 10.3390/ijerph20053894. PMID: 36900902; PMCID: PMC10002140.

[13] Fu Q, Chen Y, Li Z, Jing Q, Hu C, Liu H, Bao J, Hong Y, Shi T, Li K, Zou H, Song Y, Wang H, Wang X, Wang Y, Liu J, Liu H, Chen S, Chen R, Zhang M, Zhao J, Xiang J, Liu B, Jia J, Wu H, Zhao Y, Wan L, Xiong X. A deep learning algorithm for detection of oral cavity squamous cell carcinoma from photographic images: A retrospective study. EClinicalMedicine. 2020 Sep 23;27:100558. doi: 10.1016/j.eclinm.2020.100558. PMID: 33150326; PMCID: PMC7599313.

[14] Khanagar SB, Al-Ehaideb A, Maganur PC, Vishwanathaiah S, Patil S, Baeshen HA, Sarode SC, Bhandi S. Developments, application, and performance of artificial intelligence in dentistry - A systematic review. J Dent Sci. 2021 Jan;16(1):508-522. doi: 10.1016/j.jds.2020.06.019. Epub 2020 Jun 30. PMID: 33384840; PMCID: PMC7770297.

[15] Yu AC, Mohajer B, Eng J. External Validation of Deep Learning Algorithms for Radiologic Diagnosis: A Systematic Review. RadiolArtifIntell. 2022 May 4;4(3):e210064. doi: 10.1148/ryai.210064. PMID: 35652114; PMCID: PMC9152694.