Oral Diseases

WILEY

ORAL DISEASES

REVIEW ARTICLE OPEN ACCESS

Automated Detection of Oral Malignant Lesions Using Deep Learning: Scoping Review and Meta-Analysis

Olga Di Fede[1] | Gaetano La Mantia[1,2,3] | Marco Parola[4] | Laura Maniscalco[5] | Domenica Matranga[5] | Pietro Tozzo[6] | Giuseppina Campisi[2,7] | Mario G. C. A. Cimino[4]

[1]Department of Precision Medicine in Medical, Surgical and Critical Care (Me.Pre.C.C.), University of Palermo, Palermo, Italy | [2]Unit of Oral Medicine and Dentistry for Fragile Patients, Department of Rehabilitation, Fragility, and Continuity of Care, University Hospital Palermo, Palermo, Italy | [3]Department of Biomedical and Dental Sciences and Morphofunctional Imaging, University of Messina, Messina, Italy | [4]Department of Information Engineering, University of Pisa, Pisa, Italy | [5]Department of Health Promotion, Mother and Child Care, Internal Medicine and Medical Specialties, University of Palermo, Palermo, Italy | [6]Unit of Stomatology, Ospedali Riuniti "Villa Sofia-Cervello" of Palermo, Palermo, Italy | [7]Department of Biomedicine, Neuroscience and Advanced Diagnostics (BIND), University of Palermo, Palermo, Italy

Correspondence: Giuseppina Campisi (giuseppina.campisi@policlinico.pa.it)

Received: 26 July 2024 | **Revised:** 11 September 2024 | **Accepted:** 17 October 2024

Keywords: automatic detection | convolutional neural networks | deep learning | meta-analysis | oral cancer | oral lesion classification

ABSTRACT

Objective: Oral diseases, specifically malignant lesions, are serious global health concerns requiring early diagnosis for effective treatment. In recent years, deep learning (DL) has emerged as a powerful tool for the automated detection and classification of oral lesions. This research, by conducting a scoping review and meta-analysis, aims to provide an overview of the progress and achievements in the field of automated detection of oral lesions using DL.

Materials and Methods: A scoping review was conducted to identify relevant studies published in the last 5 years (2018–2023). A comprehensive search was conducted using several electronic databases, including PubMed, Web of Science, and Scopus. Two reviewers independently assessed the studies for eligibility and extracted data using a standardized form, and a meta-analysis was conducted to synthesize the findings.

Results: Fourteen studies utilizing various DL algorithms were identified and included for the detection and classification of oral lesions from clinical images. Among these, three were included in the meta-analysis. The estimated pooled sensitivity and specificity were 0.86 (95% confidence interval [CI] = 0.80–0.91) and 0.67 (95% CI = 0.58–0.75), respectively.

Conclusions: The results of meta-analysis indicate that DL algorithms improve the diagnosis of oral lesions. Future research should develop validated algorithms for automated diagnosis.

Trial Registration: Open Science Framework (https://osf.io/4n8sm)

1 | Introduction

Oral lesions remain a substantial global health issue affecting individuals of all age groups. In particular, potentially malignant disorders of the oral cavity (OPMD) encompass a spectrum of conditions that arise within the oral cavity and have the capacity to progress to a malignant phenotype. Timely identification and accurate diagnosis of oral lesions are crucial in preventing their progression to cancer. Notably, patients with oral cancer exhibit high mortality rates, with an estimated 50% experiencing fatal outcomes within 5 years of their initial diagnosis (Kumari, Debta, and Dixit 2022; González-Moles, Aguilar-Ruiz, and Ramos-García 2022).

However, accurate diagnosis of oral lesions remains challenging because of their wide range of variations, diverse presentations,

https://doi.org/10.1111/odi.15188

and overlapping characteristics. Ongoing efforts aim to overcome these challenges, enhance diagnostic accuracy, and develop effective interventions and management strategies for patients with oral lesions (Romano et al. 2021; González-Moles, Aguilar-Ruiz, and Ramos-García 2022).

Recent technological advancements have facilitated the development of innovative strategies and approaches to enhance early diagnosis of oral lesions, particularly for nonspecialized oral medicine practitioners (Mascitti et al. 2018). Scientific literature has proposed the use of systems that assist clinicians in identifying and accurately diagnosing lesions using dedicated apps. A notable example is the DoctOral app, which integrates algorithms specifically designed to guide clinicians in making suspected diagnoses and facilitates consultations with expert oral medicine physicians via teledentistry (Di Fede, La Mantia, et al. 2023; Di Fede, Panzarella, et al. 2023).

Another application, the mobile mouth screening anywhere (MeMoSA), serves as a communication platform that streamlines the collaboration between dentists and specialists, facilitating the assessment of management decisions concerning oral lesions and promoting effective teamwork (Dailah 2022).

In addition to these platforms, deep learning (DL) algorithms, a subset of artificial intelligence (AI), have shown immense potential for enhancing the accuracy and efficiency of oral lesion diagnosis. These algorithms can learn from extensive image and data datasets to identify patterns and features that aid in lesion detection and classification (Patil et al. 2022). The use of DL algorithms for the detection and classification of oral lesions has been extensively investigated, demonstrating their potential to improve diagnostic accuracy (Tanriver, Soluk Tekkesin, and Ergen 2021; Keser et al. 2023; Dixit, Kumar, and Srinivasan 2023; Ünsal et al. 2023).

Moreover, the application of AI in the medical field has spurred extensive research focused on developing DL systems for automating the detection and classification of oral lesions. These systems leverage images acquired using various diagnostic techniques, such as laser confocal endomicroscopy, autofluorescence imaging, hyperspectral imaging, optical coherence tomography (OCT), and clinical imaging (Aubreville et al. 2017; Fu et al. 2020; Duran-Sierra et al. 2021; Yang et al. 2023). The widespread availability of cameras, both as stand-alone devices and integrated features in smartphones, has made photographic documentation of oral lesions more accessible in medical environments (Hunt, Ruiz, and Pogue 2021; Fonseca et al. 2022; Di Fede, Panzarella, et al. 2023). Consequently, numerous tools have been developed, and various studies have been conducted to fully harness these resources and improve the detection and classification of oral lesions using advanced machine-learning techniques (Aubreville et al. 2017; Tanriver, Soluk Tekkesin, and Ergen 2021; Chu et al. 2021; Jubair et al. 2022; Yang et al. 2023).

This study aimed to provide, for the first time, a scoping review with comprehensive insight into the progress and achievements in the field of automated detection and classification of oral lesions through the application of DL to clinical images acquired via digital camera. This will be accomplished through a meta-analysis evaluating the performance metrics of specificity and sensitivity, which assess the effectiveness and accuracy of automated systems.

2 | Materials and Methods

This is the first study to employ a scoping review methodology to identify pertinent studies published within the past 5 years (2018–2023). A comprehensive search was conducted across multiple electronic databases, including PubMed, Web of Science, and Scopus, using specific keywords such as "oral lesions," "oral squamous cell carcinoma (OSCC)," "oral potentially malignant disorders," "deep learning," "neural networks," "convolutional neural networks," "machine learning," "artificial intelligence," and "automated diagnosis." Two independent reviewers (GLM and MP) screened the titles and abstracts of the identified studies to determine their eligibility. The full texts of the eligible studies were thoroughly examined, and the data were extracted using a standardized form.

During the initial search, 377 articles published between January 2018 and March 2023 were identified. This scoping review adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis-Scoping Review (PRISMA-ScR) guidelines and the Arksey and O'Malley's five-stage framework for identifying available evidence (Figure 1). The review encompassed five iterative stages: (i) formulating the research question, (ii) identifying relevant studies, (iii) selecting appropriate studies, (iv) organizing and analyzing the gathered data, and (v) summarizing the obtained results.

2.1 | Eligibility Criteria

The inclusion criteria were as follows: (i) articles published between 2018 and 2023; (ii) studies focusing on oral imaging lesions; and (iii) studies that included statistical analyses.

The lack of statistical analyses can undermine the validity of the results, making it challenging to assess the significance and generalizability of the conclusions. Therefore, studies that did not include adequate statistical analysis were excluded to ensure that the review was based on robust and reliable scientific evidence.

Specifically, precision, specificity, sensitivity, F1 score, area under the curve (AUC), and recall score were used as statistical measures to evaluate the effectiveness and accuracy of the DL classification and detection architectures.

2.2 | Study Selection

Studies were identified through an electronic search of scientific articles in different biomedical databases (Scopus, PubMed, and MEDLINE). To minimize bias, publications were examined individually by two reviewers (G.L.M. and M.P.). Any disagreements were handled through mutual conversation among the authors.

2.3 | Search Strategy

The screening procedure was guided by the predefined inclusion and exclusion criteria listed in Table 1. The following

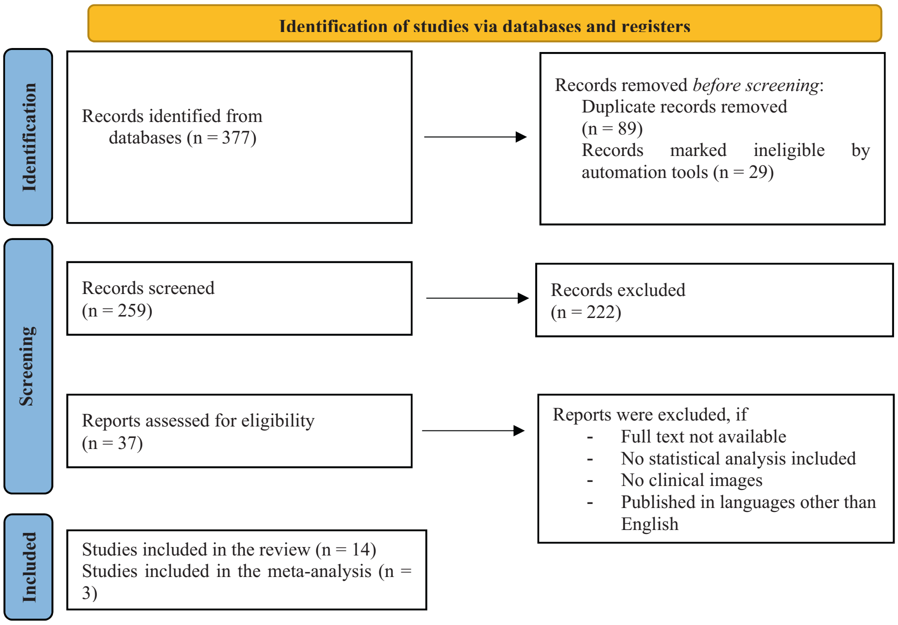

FIGURE 1 | PRISMA flow diagram for scoping reviews.

TABLE 1 | Inclusion and exclusion criteria.

Inclusion	Exclusion
Published from January 2018	Lack of statistical analysis
Primary, peer-reviewed articles Availability of full text	Languages other than English
English language	Systematic review

search terms were used separately and in combination: The search keywords included “oral lesions,”: “OSCC,” “oral potentially malignant disorders,” “deep learning,” “neural networks,” “convolutional neural networks,” “machine learning,” “artificial intelligence,” “automated diagnosis,” and free text. Full-text screening was performed only by the first author, which is typical when conducting scoping reviews. After removing duplicates, the articles were further scrutinized to assess their eligibility.

2.4 | Meta-Analysis

A meta-analysis was conducted to obtain a pooled estimate of sensitivity and specificity to quantify the effectiveness of the DL-based classification of oral lesions. The STATA module—“metandi”—was used to obtain these pooled estimates, incorporating the confusion matrix, the sequence of the number of true positives, false positives, false negatives, and true negatives. Specifically, the confusion matrix is a 2×2 table that contains the four outcomes produced by a binary classifier. A typical confusion matrix displays the actual classification of the presence/absence of a specific disease in rows and the predicted classification of the presence/absence of a specific disease in columns. In this analysis, the predicted classification was generated using DL applied to clinical images. This classification produces four outcomes: true positive, true negative, false positive, and false negative, which enables the calculation of various measures, such as specificity and sensitivity. If multiple matrices were available from the same article, each matrix was assessed individually.

3 | Results

A total of 14 studies were identified and included in the scoping review. These studies employed various DL algorithms, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and deep belief networks (DBNs), to detect and classify oral lesions.

A detailed analysis of the 14 studies revealed notable heterogeneity in the choice of DL architectures, study objectives, and dataset sizes used. Architectures range from ResNet-101, EfficientNet-b4, DenseNet-121, and VGG19 to HRNet-W18, each adopted in specific research contexts. This diversity reflects the ongoing quest for optimal solutions to specific challenges of oral pathologies.

Additionally, the focus of pathologies varied across studies, with some studies exclusively focusing on oral cancer, whereas others encompassed a broader range of oral disorders, including potentially malignant disorders (Al Duhayyim et al. 2023; Welikala et al. 2020; Lin et al. 2021; Muqeet et al. 2022; Xue et al. 2022; Shamim et al. 2022; Jubair et al. 2022; Warin, Limprasert, Suebnukarn, Jinaporntham, and Jantana 2022; Marzouk et al. 2022; Warin, Limprasert, Suebnukarn, Jinaporntham, Jantana, Vicharueang 2022; Tobias et al. 2022; Liyanage et al. 2023). This suggests that the flexibility of deep neural networks can be harnessed to address a wide array of pathological conditions (Table 2).

The analysis of 14 articles revealed heterogeneous data regarding the number of images analyzed, types of lesions considered (non-neoplastic, premalignant, and malignant), proposed

TABLE 2 | Summary of studies from 2018 to 2023 on automated detection and classification of oral lesions from clinical image using deep learning.

Article (Study design)	Authors and year	Sample size	Deep learning architectures	Statistical performance outcomes*
Automated Detection and Classification of Oral Lesions Using Deep Learning for Early Detection of Oral Cancer (Rapid review)	Welikala et al. (2020)	2155 images 1744 training images 207 validation images 204 testing images 1433 annotated images	ResNet-101	Binary (lesion/no lesion) Prec: 84.77 Recall: 89.51 F1: 87.07 Binary (referral/no referral) Prec: 67.15 Recall: 93.88 F1: 78.30 Multiclass (macro-averaged) Prec: 52.13 Recall: 49.11 F1 score: 50.57
A Deep Learning Algorithm for Detection of Oral Cavity Squamous Cell Carcinoma from Photographic Images: A Retrospective Study (Retrospective study)	Fu et al. (2020)	Initial data 44,409 images Algorithm development 5575 Internal validation dataset = 401 Secondary analysis = 170 External validation dataset = 420 Clinical validation dataset = 666	DCNN	Internal validation dataset: AUC = 95 SN = 94.9 SP = 88.7 AC = 91.5 Secondary analysis on Internal validation dataset: AUC = 99 SN = 97.4 SP = 93.5 AC = 95.3 External validation dataset: AUC = 93.5 SN = 89.6 SP = 80.6 AC = 84.1 Clinical validation dataset: AUC = 0.97 SN = 91 SP = 93.3 AC = 92.3
Automatic Detection of Oral Cancer in Smartphone-based Images Using Deep Learning for Early Diagnosis (Article)	Lin et al. (2021)	455 images 228 health mucosa 76 aphthous ulcers 69 low-risk OPMDs 52 high-risk OPMDs 30 cancer	HRNet-W18	Multiclass Prec: 84.3 Sens: 83 Spec: 96.6 F1 score: 83.6

(Continues)

TABLE 2 | (Continued)

Article (Study design)	Authors and year	Sample size	Deep learning architectures	Statistical performance outcomes*
Automated Detection and Classification of Oral Lesions Using Deep Learning to Detect Oral Potentially Malignant Disorders (Article)	Tanriver, Soluk Tekkesin, and Ergen (2021)	684 images 552 training images 63 validating images 69 testing images	EfficientNet-b4 Inception-v4 DenseNet-161 ResNeXt-101_32x8d ResNet-152 EfficientNet-b3	Multiclass (Benign/ OPMDs/OSCC) EfficientNet-b4 (Better results) Prec: 86.9 Recall: 85.5 F1 score: 85.8
Performance of Deep Convolutional Neural Network for Classification and Detection of Oral Potentially Malignant Disorders in Photographic Images (Clinical paper)	Warin, Limprasert, Suebnukarn, Jinaporntham, and Jantana (2022)	600 images (including 300 OPMDs and 300 normal oral mucosa) 420 training images 60 validating images 120 testing images	DenseNet-121 ResNet-50	DenseNet: 121 Prec: 91 Sens: 100 Spec: 90 F1 score: 95 Recall: 89.51 AUC of the ROC curve: 95 ResNet-50 Prec: 92 Sens: 98.37 Spec: 91.67 F1 score: 95 Recall: 98 AUC of the ROC curve: 95.03
Automated Oral Cancer Detection using Deep Learning-based Technique (Article)	Muqeet et al. (2022)	131 images 87 OSCC 44 NO-OSCC	VGG19 InceptionV3 Xception	VGG-19 Prec: 82 Recall: 100 F1 score: 90 Acc: 85.19 InceptionNet-V3 Prec: 83.33 Recall: 100 F1 score: 91 Acc: 88.89 Xception Prec: 94.44 Recall: 100 F1 score: 97 Acc: 96.30
Deep Transfer Learning Driven Oral Cancer Detection and Classification Model (Article)	Marzouk et al. (2022)	131 images 87 OSCC 44 NO-OSCC	AIDTL-OCCM	AIDTL-OCCM Prec: 89.05 Recall: 88.60 Acc: 90.08 F1 score: 88.8

(Continues)

TABLE 2 | (Continued)

Article (Study design)	Authors and year	Sample size	Deep learning architectures	Statistical performance outcomes*
Automated Detection of Oral Pre-Cancerous Tongue Lesions Using Deep Learning for Early Diagnosis of Oral Cavity Cancer (Article)	Shamim et al. (2022)	300 images 240 training images 60 validation images	AlexNet GoogLeNet Vgg19 Inceptionv3 ResNet50 SqueezeNet	(Binary Classification) AlexNet Acc: 93 ± 0.06 Sens: 88 Spec: 94 GoogLeNet Acc: 0·93 ± 0.02 Sens: 80 Spec: 88 ResNet50 Acc: 90 ± 0.04 Sens: 84 Spec: 96 Vgg19 Acc: 98 ± 0.04 Sens: 89 Spec: 97 Inceptionv3 Acc: 93 ± 0.03 Sens: 83 Spec: 88 SqueezeNet Acc: 93 ± 0.09 Sens 85 Spec: 96
A Novel Lightweight Deep Convolutional Neural Network for Early Detection of Oral Cancer (Article)	Jubair et al. (2022)	716 images	EfficientNet-B0	Multiclass Acc: 85 Prec: 84.3 Sens: 86.7 Spec: 84.5 F1: 83.7
AI-based Analysis of Oral Lesions Using Novel Deep Convolutional Neural Networks for Early Detection of Oral Cancer (Article)	Warin, Limprasert, Suebnukarn, Jinaporntham, Jantana, Vicharueang (2022)	980 images 365 OSCC 315 OPMDs 300 non-pathological images	DenseNet169 ResNet-101 SqueezeNet Swin-S	Multiclass OSCC/ OPMDs DenseNet-169 (Better results) OSCC–OPMDs Prec: 98–95 Sens: 99–95 Spec: 99–97 F1: 98–95 AUC: 100–98

(Continues)

TABLE 2 | (Continued)

Article (Study design)	Authors and year	Sample size	Deep learning architectures	Statistical performance outcomes*
Automatic Detection of Oral Lesion Measurement Ruler Toward Computer-Aided Image-Based Oral Cancer Screening (Article)	Xue et al. (2022)	7248 images	ImageNet ResNeSt ViT	ImageNet Sens: 33 Spec: 100 Prec: 100 F1: 49.6 Acc: 89 ResNeSt Sens:100 Spec: 99.6 Prec: 97.9 F1: 98.9 Acc: 99.6 ViT Sens: 100 Spec: 99.8 Prec: 98.9 F1: 99.5 Acc: 99.8
Artificial Intelligence for Oral Cancer Diagnosis: What are the Possibilities? (Article)	Tobias et al. (2022)	1636 images	ResNet	Recall: 78.5 Acc: 78 Prec: 77.1 F1: 77
Sailfish Optimization with Deep Learning Based Oral Cancer Classification Model (Article)	Al Duhayyim et al. (2023)	131 images 87 OSCC 44 NO-OSCC	CADOC-SFOFC	Acc: 98.11 Prec: 98.72 Recall: 96.67 F1 score: 97.63
Malignant and Non-Malignant Oral Lesions Classification and Diagnosis with Deep Neural Networks (Article)	Liyanage et al. (2023)	113 images non-neoplastic lesions 107 images benign neoplasms 122 images premalignant/ malignant lesions	MobileNetV3 EfficientNetV2	MobileNetV3: Acc: 71 Prec: 61 Recall: 64 F1 score:62 AUC: 88 EfficientNetV2: Acc: 70 Prec: 64 Recall: 64 F1 score: 64 AUC: 88

Abbreviations: AUC, area under the curve; DCNN, deep convolutional neural networks; FN, false negative; FP, false positive; OPMD, potentially malignant disorders of the oral cavity; OSCC, oral squamous cell carcinoma; TN, true negative; TP, true positive.
*Legend of statistical performance outcomes: Accuracy = (TP + TN)/(P + N); Precision = TP/(TP + FP); Recall = TP/(TP + FN); F1 score = 2 × [(Precision – Recall)/(Precision + Recall)]; Sensitivity = TP/(TP + FN); Specificity = TN/(TN + FP).

deep-learning architectures, and their respective efficiency values. These diverse results underscore the need for comparative evaluation to identify the most effective system for diagnosing lesions using clinical images. However, the extensive heterogeneity of values and informational gaps limited the implementation of a meta-analysis, which included only three studies (Welikala et al. 2020; Tobias et al. 2022; Liyanage et al. 2023) that provided the confusion matrices necessary for meta-analysis. Among these studies, two provided confusion matrices for two different endpoints (Welikala et al. 2020; Liyanage et al. 2023), resulting in five confusion matrices. The pooled estimates were derived using a two-level mixed logistic regression model, with independent binomial distributions for true positives and true negatives conditional on the sensitivity and specificity in each study, and a bivariate normal model for the logit transforms of sensitivity and specificity between studies. The results are shown as sensitivity (Se) and specificity (Sp) with corresponding 95% confidence intervals (CIs).

3.1 | Results of Meta-Analysis

The meta-analysis of the five confusion matrices revealed that sensitivity ranged between 0.79 (95% CI = 0.66–0.88) (Tobias et al. 2022) and 0.94 (95% CI = 0.87–0.98) (Welikala et al. 2020), and specificity ranged between 0.58 (95% CI = 0.48–0.67) (Welikala et al. 2020) and 0.79 (95% CI = 0.67–0.89) (Tobias et al. 2022) (Table 3). The estimated pooled sensitivity was 0.86 (95% CI = 0.80–0.91) and specificity was 0.67 (95% CI = 0.58–0.75), which accounted for the varying precision of estimates, as indicated by large error margins in some 95% CIs.

4 | Discussion

Early identification of oral cavity lesions, particularly those indicative of potentially malignant disorders and oral cancer, is crucial for enhancing survival rates and reducing the associated morbidity. However, delays in clinical evaluation are common, particularly among physicians lacking expertise in oral medicine, often leading to substantial delays between the onset of initial symptoms and diagnosis (Scott, Grunfeld, and McGurk 2006; Mauceri et al. 2022).

Despite the extensive use of radiographic imaging techniques such as magnetic resonance imaging and computed tomography to assess the size and extent of oral cancer before treatment initiation, these methods have limitations in sensitivity, which limit their ability to effectively distinguish neoplastic lesions. To address this challenge, additional clinical imaging techniques, such as autofluorescence and OCT, have been introduced (Camalan et al. 2021).

However, the widespread adoption of these methodologies is often hindered by high costs and the requirement for highly specialized personnel. The integration of AI in healthcare has demonstrated substantial reductions in cost and optimized efficiency. Recently, AI-based technologies have revolutionized the early diagnosis of oral lesions through accurate analysis of photographic images (Mintz and Brodie 2019).

Our study aimed to conduct, for the first time, a comprehensive scoping review to investigate the advancements in AI technology for clinical image analysis of oral cavity lesions acquired through cameras by searching and selecting recent articles published in various relevant databases. Fourteen articles were identified, and the limited number was compensated by the relevance of the number of images available for each selected study. Furthermore, three of these studies allowed the execution of a meta-analysis using confusion matrices for the first time, which enabled the aggregation and synthesis of results to provide an overall view of the performance of various AI models applied for the detection and classification of clinical images of oral lesions.

Diagnostic accuracy values, including sensitivity and specificity, were validated through a series of investigations that successfully implemented AI analysis on diverse images. Although photographic imaging offers a quick, convenient and accessible approach, variability in image quality remains a major limitation. Factors such as the type of camera, ambient lighting, and resolution can influence image quality, potentially limiting the accuracy. In particular, the oral cavity, a complex and three-dimensional anatomical context encompassing the lips, teeth, and oral mucosa, presents unique challenges for accurate image capture (Fu et al. 2020).

Recently, a smartphone-based device with a probe for easy access to the mouth was introduced to enhance image quality (Uthoff et al. 2018). The need for a method that allows the standardization of image acquisition is currently lacking, and several studies have proposed the use of smartphones for rapid image capture (Benjumea et al. 2020; Dailah 2022; Di Fede, Panzarella, et al. 2023).

A recent systematic review explored the use of AI systems for the automatic recognition of potentially malignant lesions and OSCC from clinical images acquired with various devices, such as autofluorescence and OCT. The results aligned with our findings based on clinical images acquired with digital cameras. However, these tools entail higher acquisition costs than standard digital cameras or smartphones with cameras (Li et al. 2024).

TABLE 3 | Results of the meta-analysis of the three included studies analyzing the automated detection and classification of oral clinical images lesions using deep learning.

Article		TP	FP	TN	FN	Se	95% CI	Sp	95% CI
Welikala et al. (2020)	Binary (lesion/no lesion)	128	23	38	15	0.9	0.83–0.94	0.62	0.49–0.74
Welikala et al. (2020)	Binary (referral/no referral)	92	45	61	6	0.94	0.87–0.98	0.58	0.48–0.67
Liyanage et al. (2023)	Multiclass MobileNetV3	36	8	14	9	0.8	0.65–0.90	0.64	0.41–0.83
Liyanage et al. (2023)	Multiclass EfficientNetV2	37	8	14	8	0.82	0.68–0.92	0.64	0.41–0.83
Tobias et al. (2022)	Binary (OSCC-OPMDs/ health mucosa)	44	13	50	12	0.79	0.66–0.88	0.79	0.67–0.89
	Overall	337	97	177	50	0.86	0.80–0.91	0.67	0.58–0.75

Abbreviations: 95% CI, 95% confidence interval; FN, false negative; FP, false positive; OPMD, potentially malignant disorders of the oral cavity; OSCC, oral squamous cell carcinoma; Se, sensitivity; Sp, specificity; TN, true negative; TP, true positive.

AI-assisted diagnosis of clinical images acquired through smartphones offers broad global accessibility and enables precise and efficient screening through AI analysis of a large number of images. However, it is crucial to ensure that these practices comply with the privacy regulations, such as the General Data Protection Regulation (GDPR) in Europe and the Health Insurance Portability and Accountability Act (HIPAA) in the USA (Di Fede, La Mantia, et al. 2023).

Our findings highlight the significant sensitivity of AI-assisted diagnosis in analyzing photographic images, emphasizing its usefulness in facilitating quick clinical decision-making regarding further examinations and treatments. In addition to providing high-precision diagnostic support, this approach could also contribute to cost containment in the context of public health by preventing overload and optimizing resource use (Fu et al. 2020).

The distinctive ability of AI to accurately identify precancerous lesions with a sensitivity exceeding 90% compared to normal tissues confirms the exceptional accuracy of this screening method. This outcome not only reinforces confidence in AI-assisted technology but also suggests considerable potential for improving clinical decision making. Advanced precision in discriminating pathological conditions could have a substantial impact on the timely and targeted management of diseases, making a tangible contribution to public health. Numerous studies have demonstrated that analyzed AI systems have higher sensitivity than specificity. For instance, Welikala et al. (Welikala et al. 2020) reported a sensitivity of 89.51% and a specificity of 67.15% for early detection of oral cancer. This trend remained consistent across other studies, such as that of Lin et al. (Lin et al. 2021), who reported a sensitivity of 83% and a specificity of 96.6%.

Despite these variations, our aggregate analysis revealed that, on average, these models exhibited an estimated sensitivity of 86% and a specificity of 67%. Although the balance between sensitivity and specificity is generally positive, the variations underscore the importance of considering the specific contexts in which these models are applied. Recently, a study proposed an intriguing idea, suggesting the use of an ensemble of architectures to optimize both learning and lesion detection. This comprehensive strategy leverages the specific strengths of individual models to achieve superior performance than those of single models. The proposed architecture was evaluated using a real-world dataset created by expert clinicians who manually labeled individual photographic images, thereby generating a reference dataset. Comparative analysis results clearly demonstrated that the ensemble detection model outperformed the single models, exceeding the average value of the map@50 metric by 24% and the value of the map@95–50 metric by 44% (Parola et al. 2023).

The results of this scoping review should be considered in light of the advantages and limitations of AI. Machine learning and deep learning approaches enable systems to learn from data without explicit programming and can handle heterogeneous data sources. However, addressing all potential biases is crucial for ensuring the reliability and accuracy of study results (Barbati et al. 2023). In this scoping review, the sizes of the training datasets varied considerably, ranging from studies with a few hundred images to those with thousands of images. Dataset size is a critical factor that can significantly influence model performance and introduce external validation bias. These data indicate the need to balance the model performance in terms of sensitivity and specificity, considering the importance of both indicators in the early diagnosis of oral pathologies. The analyzed studies demonstrated promising results, with high accuracy rates in detecting and classifying oral lesions using deep learning algorithms.

While keeping the debate open regarding the potential replacement of experts with AI, it is hoped that the integration of AI in oral cancer diagnosis can effectively contribute to significant reduction in mortality and morbidity, especially in low- and middle-income countries where healthcare systems face substantial gaps (Ciecierski-Holmes et al. 2022).

The integration of AI in the early diagnosis of oral cancer offers a promising avenue for reducing cancer-related morbidity and mortality (Celentano and Cirillo 2024). Identifying precancerous lesions is a key strategy to achieve this goal (Warnakulasuriya et al. 2021; Celentano and Cirillo 2024). However, despite ongoing efforts, global mortality rates of oral cancer have remained unchanged over the past 30 years (Byers et al. 2016). Epidemiological evidence has revealed significant disparities in oral cancer incidence and survival, linked to sociodemographic and geographical variables, likely attributed to differences in access to healthcare and patterns of risk factor exposure (Celentano and Cirillo 2024). This also underscores the urgent need for implementing remote consultation systems through telemedicine. AI-powered telemedicine can substantially enhance the access to care for patients residing in remote areas, with limited access to healthcare facilities, particularly those with limited healthcare resources (Schroeder et al. 2024). This approach helps alleviate the workload of healthcare professionals, improves the quality of care, and reduces healthcare expenditure for both healthcare providers and patients (Ribeiro-Rotta et al. 2022; Ben Dor et al. 2024). To strengthen these benefits, it is crucial to overcome political and economic barriers, promote equitable access to care, and ensure that even the most disadvantaged populations benefit from advanced diagnostic services (Celentano and Cirillo 2024). Large-scale collection of image data is essential for refining the accuracy of AI-assisted analysis from a clinical perspective. This approach offers a promising avenue to elevate diagnostic standards and enhance the overall effectiveness of medical interventions in less-developed healthcare settings.

5 | Limitations

Despite being the first meta-analysis to assess the effectiveness of various AI architectures for lesion detection using AI-based image analysis, our study had several limitations that needs to be acknowledged. First, despite analyzing the applications of diverse AI architectures with the same imaging tool, variations in the quality of the devices used in each study and differences in techniques can impact diagnostic accuracy. To date, no studies in the literature have systematically compared the quality and imaging techniques used in different studies applying AI systems for lesion recognition and performance evaluation. Second, our interpretation of the results was constrained by the non-availability of prospective studies comparing conventional examinations with AI- based diagnosis through clinical imaging. Addressing these challenges in various clinical contexts is crucial for advancing AI-assisted healthcare. To improve reliability and generalizability of the findings, future research with larger and more diverse sample sizes is recommended to obtain more accurate and robust estimates.

6 | Conclusions

Early detection of oral cancer is crucial; however, delays in the diagnosis of oral mucosal lesions persist. AI exhibits high sensitivity and low specificity for analyzing photographic images, overcoming the limitations of conventional techniques. To foster an innovative future, the adoption of portable devices with advanced AI technologies could enable precise diagnosis and easy access to diagnostic approaches, thereby enhancing treatment effectiveness and patient prognosis.

All authors have provided their final approval regarding the study contents and consent to be accountable for all aspects of this work. All the authors have read and agreed to the published version of the manuscript. All the authors have read and agreed to the published version of the manuscript.

Author Contributions

Olga Di Fede: conceptualization, methodology, investigation, supervision, writing – original draft, writing – review and editing, validation. **Gaetano La Mantia:** conceptualization, writing – original draft, investigation, methodology. **Marco Parola:** data curation, investigation, visualization, writing – original draft. **Laura Maniscalco:** investigation, methodology, formal analysis, software. **Domenica Matranga:** investigation, validation, visualization, software, writing – original draft. **Pietro Tozzo:** data curation, investigation. **Giuseppina Campisi:** conceptualization, validation, supervision, visualization, writing – review and editing, methodology, investigation. **Mario G. C. A. Cimino:** writing – review and editing, supervision, validation, methodology, conceptualization.

Acknowledgments

This study was partially supported by: (i) the European Commission under the NextGenerationEU program, (ii) Extended partnership-PNRR PE1—“FAIR—Future Artificial Intelligence Research”—Spoke 1 “Human-centered AI”; (iii) the Italian Ministry of Education and Research (MIUR) in the framework of the FoReLab project (Departments of Excellence), in the framework of the “Reasoning” project, PRIN 2020 LS Programme, Project number 2493 04-11-2021, and in the framework of the project “OCAX—Oral CAncer eXplained by DL-enhanced case-based classification” PRIN 2022 code P2022KMWX3. Received partial funding from the European Commission under the NextGeneration EU program, PNRR—M4 C2, Investment 1.5 Creating and strengthening of “innovation ecosystems,” building “territorial R&D leaders,” project “THE: Tuscany Health Ecosystem,” Spoke 6 “Precision Medicine and Personalized Healthcare”. The article processing charge was funded by the University of Palermo (Piano straordinario VQR 2020-2024: Misura A to ODF and GC.

Consent

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Al Duhayyim, M., A. A. Malibari, S. Dhahbi, et al. 2023. “Sailfish Optimization with Deep Learning Based Oral Cancer Classification Model.” *Computer Systems Science and Engineering* 45, no. 1: 753–767.

Aubreville, M., C. Knipfer, N. Oetter, et al. 2017. “Automatic Classification of Cancerous Tissue in Laserendomicroscopy Images of the Oral Cavity Using Deep Learning.” *Scientific Reports* 7: 1–10.

Barbati, G., P. Pasqualetti, D. Matranga, et al. 2023. “Study Design and Research Protocol for Diagnostic or Prognostic Studies in the Age of Artificial Intelligence: A Biostatistician's Perspective.” *Epidemiology, Biostatistics and Public Health* 18: 12.

Ben Dor, B., A. Villa, C. Hayes, E. Alpert, D. S. Shepard, and S. T. Sonis. 2024. “Financial Burden of Dental Care Among Patients With Head and Neck Cancer.” *JAMA Otolaryngology. Head & Neck Surgery* 150: 811–818.

Benjumea, J., J. Ropero, O. Rivera-Romero, E. Dorronzoro-Zubiete, and A. Carrasco. 2020. “Privacy Assessment in Mobile Health Apps: Scoping Review.” *JMIR mHealth and uHealth* 8: e18868.

Byers, T., R. C. Wender, A. Jemal, A. M. Baskies, E. E. Ward, and O. W. Brawley. 2016. “The American Cancer Society Challenge Goal to Reduce US Cancer Mortality by 50% Between 1990 and 2015: Results and Reflections.” *CA: A Cancer Journal for Clinicians* 66: 359–369.

Camalan, S., H. Mahmood, H. Binol, et al. 2021. “Convolutional Neural Network-Based Clinical Predictors of Oral Dysplasia: Class Activation Map Analysis of Deep Learning Results.” *Cancers* 13: 1291.

Celentano, A., and N. Cirillo. 2024. “Diseases With Oral Malignant Potential: Need for Change to Inform Research, Policy, and Practice.” *Journal of Oral Pathology & Medicine* 53: 495–501.

Chu, C. S., N. P. Lee, J. W. K. Ho, S. W. Choi, and P. J. Thomson. 2021. “Deep Learning for Clinical Image Analyses in Oral Squamous Cell Carcinoma: A Review.” *JAMA Otolaryngology. Head & Neck Surgery* 147: 893–900.

Ciecierski-Holmes, T., R. Singh, M. Axt, S. Brenner, and S. Barteit. 2022. “Artificial Intelligence for Strengthening Healthcare Systems in Low- and Middle-Income Countries: A Systematic Scoping Review.” *NPJ Digital Medicine* 5: 1–13.

Dailah, H. G. 2022. “Mobile Health (mHealth) Technology in Early Detection and Diagnosis of Oral Cancer—A Scoping Review of the Current Scenario and Feasibility.” *Journal of Healthcare Engineering* 2022: 1–11.

Di Fede, O., G. La Mantia, M. G. C. A. Cimino, and G. Campisi. 2023a. “Protection of Patient Data in Digital Oral and General Health Care: A Scoping Review With Respect to the Current Regulations.” *Oral* 3: 155–165.

Di Fede, O., V. Panzarella, F. Buttacavoli, G. La Mantia, and G. Campisi. 2023b. “Doctoral: A Smartphone-Based Decision Support Tool for the Early Detection of Oral Potentially Malignant Disorders.” *DIGITAL HEALTH* 9: 20552076231177141.

Dixit, S., A. Kumar, and K. Srinivasan. 2023. “A Current Review of Machine Learning and Deep Learning Models in Oral Cancer Diagnosis: Recent Technologies, Open Challenges, and Future Research Directions.” *Diagnostics* 13: 1353.

Duran-Sierra, E., S. Cheng, R. Cuenca, et al. 2021. “Machine-Learning Assisted Discrimination of Precancerous and Cancerous From Healthy Oral Tissue Based on Multispectral Autofluorescence Lifetime Imaging Endoscopy.” *Cancers* 13: 4751.

Fonseca, B. B., N. N. Perdoncini, V. C. da Silva, et al. 2022. “Telediagnosis of Oral Lesions Using Smartphone Photography.” *Oral Diseases* 28: 1573–1579.

Fu, Q., Y. Chen, Z. Li, et al. 2020. “A Deep Learning Algorithm for Detection of Oral Cavity Squamous Cell Carcinoma From Photographic Images: A Retrospective Study.” *EClinicalMedicine* 27: 100558.

González-Moles, M. Á., M. Aguilar-Ruiz, and P. Ramos-García. 2022. “Challenges in the Early Diagnosis of Oral Cancer, Evidence Gaps and Strategies for Improvement: A Scoping Review of Systematic Reviews.” *Cancers* 14: 4967.

Hunt, B., A. J. Ruiz, and B. W. Pogue. 2021. “Smartphone-Based Imaging Systems for Medical Applications: A Critical Review.” *Journal of Biomedical Optics* 26: 40902–40903.

Jubair, F., O. Al-karadsheh, D. Malamos, S. Al Mahdi, Y. Saad, and Y. Hassona. 2022. “A Novel Lightweight Deep Convolutional Neural Network for Early Detection of Oral Cancer.” *Oral Diseases* 28: 1123–1130.

Keser, G., İ. Ş. Bayrakdar, F. N. Pekiner, Ö. Çelik, and K. Orhan. 2023. “A Deep Learning Algorithm for Classification of Oral Lichen Planus Lesions From Photographic Images: A Retrospective Study.” *Journal of Stomatology Oral and Maxillofacial Surgery* 124: 101264.

Kumari, P., P. Debta, and A. Dixit. 2022. “Oral Potentially Malignant Disorders: Etiology, Pathogenesis, and Transformation Into Oral Cancer.” *Frontiers in Pharmacology* 13: 825266.

Li, J., W. Y. Kot, C. P. McGrath, B. W. A. Chan, J. W. K. Ho, and L. W. Zheng. 2024. “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.” *International Journal of Surgery* 110, no. 8: 5034–5046.

Lin, H., H. Chen, L. Weng, J. Shao, and J. Lin. 2021. “Automatic Detection of Oral Cancer in Smartphone-Based Images Using Deep Learning for Early Diagnosis.” *Journal of Biomedical Optics* 26: 086007.

Liyanage, V., M. Tao, J. S. Park, K. N. Wang, and S. Azimi. 2023. “Malignant and Non-Malignant Oral Lesions Classification and Diagnosis With Deep Neural Networks.” *Journal of Dentistry* 137: 104657.

Marzouk, R., E. Alabdulkreem, S. Dhahbi, et al. 2022. “Deep Transfer Learning Driven Oral Cancer Detection and Classification Model.” *Computers, Materials & Continua* 73: 3905–3920.

Mascitti, M., G. Orsini, V. Tosco, et al. 2018. “An Overview on Current Non-Invasive Diagnostic Devices in Oral Oncology.” *Frontiers in Physiology* 9: 1510.

Mauceri, R., M. Bazzano, M. Coppini, P. Tozzo, V. Panzarella, and G. Campisi. 2022. “Diagnostic Delay of Oral Squamous Cell Carcinoma and the Fear of Diagnosis: A Scoping Review.” *Frontiers in Psychology* 13: 1009080.

Mintz, Y., and R. Brodie. 2019. “Introduction to Artificial Intelligence in Medicine.” *Minimally Invasive Therapy & Allied Technologies* 28: 73–81. https://doi.org/10.1080/13645706.2019.1575882.

Muqeet, M. A., A. B. Mohammad, P. G. Krishna, S. H. Begum, S. Qadeer, and N. Begum. 2022. “Automated Oral Cancer Detection Using Deep Learning-Based Technique.” *8th International Conference on Signal Processing and Communication*, ICSC 2022: 294–297.

Parola, M., G. La Mantia, F. Galatolo, M. G. C. A. Cimino, G. Campisi, and O. Di Fede. 2023. “Image-Based Screening of Oral Cancer via Deep Ensemble Architecture.” *EEE Symposium Series on Computational Intelligence* (SSCI) Mexico City, Mexico, December 5–8, 2023.

Patil, S., S. Albogami, J. Hosmani, et al. 2022. “Artificial Intelligence in the Diagnosis of Oral Diseases: Applications and Pitfalls.” *Diagnostics* 12: 1029.

Ribeiro-Rotta, R. F., E. A. Rosa, V. Milani, et al. 2022. “The Cost of Oral Cancer: A Systematic Review.” *PLoS One* 17: e0266346.

Romano, A., D. Di Stasio, M. Petruzzi, et al. 2021. “Noninvasive Imaging Methods to Improve the Diagnosis of Oral Carcinoma and Its Precursors: State of the Art and Proposal of a Three-Step Diagnostic Process.” *Cancers* 13: 2864.

Schroeder, K., M. Santoro, E. P. Tranby, et al. 2024. “Teledentistry Utilization by Oral Health Professionals and Policy Considerations: A Mixed Methods Case Study.” *Journal of Public Health Dentistry*. Published ahead of print.

Scott, S. E., E. A. Grunfeld, and M. McGurk. 2006. “Patient's Delay in Oral Cancer: A Systematic Review.” *Community Dentistry and Oral Epidemiology* 34: 337–343.

Shamim, M. Z. M., S. Syed, M. Shiblee, et al. 2022. “Automated Detection of Oral Pre-Cancerous Tongue Lesions Using Deep Learning for Early Diagnosis of Oral Cavity Cancer.” *Computer Journal* 65: 91–104.

Tanriver, G., M. Soluk Tekkesin, and O. Ergen. 2021. “Automated Detection and Classification of Oral Lesions Using Deep Learning to Detect Oral Potentially Malignant Disorders.” *Cancers* 13: 2766.

Tobias, M. A. S., B. P. Nogueira, M. C. S. Santana, R. G. Pires, J. P. Papa, and P. S. S. Santos. 2022. “Artificial Intelligence for Oral Cancer Diagnosis: What Are the Possibilities?” *Oral Oncology* 134: 106117.

Ünsal, G., A. Chaurasia, N. Akkaya, et al. 2023. “Deep Convolutional Neural Network Algorithm for the Automatic Segmentation of Oral Potentially Malignant Disorders and Oral Cancers.” *Proceedings of the Institution of Mechanical Engineers. Part H* 237: 719–726.

Uthoff, R. D., B. Song, S. Sunny, et al. 2018. “Point-of-Care, Smartphone-Based, Dual-Modality, Dual-View, Oral Cancer Screening Device With Neural Network Classification for Low-Resource Communities.” *PLoS One* 13: e0207493.

Warin, K., W. Limprasert, S. Suebnukarn, S. Jinaporntham, and P. Jantana. 2022a. “Performance of Deep Convolutional Neural Network for Classification and Detection of Oral Potentially Malignant Disorders in Photographic Images.” *International Journal of Oral and Maxillofacial Surgery* 51: 699–704.

Warin, K., W. Limprasert, S. Suebnukarn, S. Jinaporntham, P. Jantana, and S. Vicharueang. 2022b. “AI-Based Analysis of Oral Lesions Using Novel Deep Convolutional Neural Networks for Early Detection of Oral Cancer.” *PLoS One* 17: e0273508.

Warnakulasuriya, S., O. Kujan, J. M. Aguirre-Urizar, et al. 2021. “Oral Potentially Malignant Disorders: A Consensus Report From an International Seminar on Nomenclature and Classification, Convened by the WHO Collaborating Centre for Oral Cancer.” *Oral Diseases* 27: 1862–1880.

Welikala, R. A., P. Remagnino, J. H. Lim, et al. 2020. “Automated Detection and Classification of Oral Lesions Using Deep Learning for Early Detection of Oral Cancer.” *IEEE Access* 8: 132677–132693.

Xue, Z., K. Yu, P. C. Pearlman, et al. 2022. “Automatic Detection of Oral Lesion Measurement Ruler Toward Computer-Aided Image-Based Oral Cancer Screening.” *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* 2022: 3218.

Yang, Z., H. Pan, J. Shang, J. Zhang, and Y. Liang. 2023. “Deep-Learning-Based Automated Identification and Visualization of Oral Cancer in Optical Coherence Tomography Images.” *Biomedicine* 11: 802.