GRADUATION PROJECT

Degree in Dentistry

CAN ARTIFICIAL INTELLIGENCE BE USEFUL AGAINST ORAL CANCER?

Madrid, academic year 2025/2026

Identification number: 169

RESUMEN

Introducción: El cáncer oral es una neoplasia frecuente de la región de cabeza y cuello y sigue siendo un problema relevante por su incidencia, el diagnóstico tardío y la escasa mejoría de la supervivencia. La inteligencia artificial (IA) ha surgido como una herramienta prometedora para apoyar el cribado, la detección precoz, el diagnóstico y la predicción del riesgo en el cáncer oral y en los trastornos orales potencialmente malignos; **Objetivos**: Realizar una revisión sistemática de la literatura para evaluar la utilidad de sistemas basados en IA en la detección precoz, el diagnóstico y la predicción pronostica del cáncer oral, comparando, cuando fue posible, su rendimiento con los métodos convencionales; **Metodología**: Se efectuó una búsqueda sistemática en Medline Complete e IEEE Xplore de estudios originales en inglés publicados en los últimos cinco años que incluyeran pacientes con cáncer oral o trastornos orales potencialmente malignos, aplicaran sistemas de IA y reportaran al menos una métrica cuantitativa. Tras un proceso de cribado independiente, se seleccionaron 20 estudios; **Resultados**: Los modelos de IA mostraron altos valores de precisión, sensibilidad, especificidad y AUC en tareas de cribado, clasificación de lesiones, segmentación y predicción del riesgo, aunque la mayoría de los estudios fueron retrospectivos, de un solo centro y sin validación externa robusta; **Conclusiones**: En consecuencia, la IA presenta un notable potencial como herramienta complementaria para mejorar el cribado, la consistencia diagnóstica y la estratificación del riesgo en cáncer oral, pero aún no puede sustituir a los métodos diagnósticos establecidos ni al juicio clínico especializado.

PALABRAS CLAVE

Odontología, cáncer oral, inteligencia artificial, diagnóstico, detección precoz.

ABSTRACT

Introduction: Oral cancer is a frequent malignancy of the head and neck and remains a major health problem due to its incidence, late-stage diagnosis and modest improvement in survival. Early detection is hindered by subtle clinical presentations and the subjectivity of visual examination. In recent years, artificial intelligence (AI), particularly machine learning and deep learning, has emerged as a promising tool to support screening, diagnosis and prognostic assessment in oral cancer and oral potentially malignant disorders; **Objectives**: To perform a systematic literature review to evaluate the usefulness of AI-based systems in the early detection, diagnosis and risk prediction of oral cancer, and to compare, when possible, their performance with conventional diagnostic methods; **Methods**: A systematic search was conducted in Medline Complete and IEEE Xplore. Original English-language studies from the last five years were included if they involved patients with oral cancer or oral potentially malignant disorders, applied AI-based systems and reported at least one quantitative performance metric. After independent screening and eligibility assessment, 20 primary studies were selected; **Results**: AI models using clinical photographs, smartphone images, histopathology, radiology and tabular clinical data showed high accuracy, sensitivity, specificity and AUC in tasks such as screening, lesion classification, segmentation and prognostic prediction. In several studies, AI systems matched or outperformed traditional grading systems or statistical models, although most were retrospective, single-center and lacked robust external validation; **Conclusions**: AI-based systems show substantial potential as complementary tools to improve screening, diagnostic consistency and risk stratification in oral cancer, but they cannot yet replace established diagnostic standards.

KEYWORDS

Odontology, oral cancer, artificial intelligence, diagnosis, early detection.

TABLE OF CONTENTS

1.	INTRODUCTION	5
2.	OBJECTIVE	9
3.	MATERIAL AND METHODS	11
4.	RESULTS	14
5.	DISCUSSION	26
6.	CONCLUSIONS	32
7.	SUSTAINABILITY	34
8.	REFERENCES	36
9.	ANNEXES	40

1. INTRODUCTION

According to the World Health Organization, cancer is defined as “a large group of diseases that can start in almost any organ or tissue of the body when abnormal cells grow uncontrollably, go beyond their usual boundaries to invade adjoining parts of the body and/or spread to other organs” (1). Among these malignancies, oral cancer (OC) involves the lips, tongue, floor of the mouth, buccal mucosa, gingiva, salivary glands, and other structures of the oral cavity (1).

Oral cancer represents the most common malignancy of the head and neck region and remains a significant global health concern. According to GLOBOCAN 2020 data, approximately 389,846 new cases and 188,438 deaths were reported worldwide in that year (2). These figures refer to estimated new cases and deaths for the year 2020, reflecting the annual global burden of the disease. Its geographical distribution is uneven, with higher prevalence observed in parts of Asia, Europe, and South America, and it affects men more frequently than women (3)

Approximately 90% of oral cancers are oral squamous cell carcinomas (OSCC), which are characterized by aggressive behavior and a marked potential for local invasion and regional or distant metastasis (2). The development of OSCC is strongly associated with well-established lifestyle risk factors, including tobacco use, alcohol consumption, betel nut chewing, and cannabis smoking. Additionally, human papillomavirus infection, increasing age, and genetic predisposition have also been implicated in its pathogenesis (4).

In many patients, OSCC develops from pre-existing oral potentially malignant disorders (OPMDs). These conditions include leukoplakia, erythroplakia, proliferative verrucous leukoplakia, oral submucous fibrosis, and oral lichen planus, among others. OPMDs are characterized by clinically detectable mucosal alterations associated with an increased risk of malignant transformation, particularly in non-homogeneous leukoplakia (5,6). For this reason, early identification and close monitoring of these lesions are essential to reduce the likelihood of progression to invasive carcinoma.

Currently, the diagnosis of oral cancer relies primarily on clinical examination followed by biopsy and histopathological analysis, which remains the gold standard. Adjunctive diagnostic tools such as radiographs, computed tomography (CT), magnetic resonance imaging, light-based

detection systems, and brush cytology may assist in evaluating tumor extension and dissemination (6). However, these approaches can present several limitations. Clinical assessment is partly dependent on the clinician's experience and may lead to inter-observer variability. Biopsy procedures are invasive and may cause discomfort for the patient, while imaging techniques can be time-consuming and costly (7).

Despite advances in treatment modalities, the overall survival rate of oral cancer has not substantially improved over recent decades (7). One of the major reasons is the frequent diagnosis at advanced stages, often due to the subtle and non-specific presentation of early lesions. Consequently, there is a clear need for complementary tools that could enhance early detection, improve diagnostic consistency, and support clinical decision-making without replacing established diagnostic standards (8).

In this context, Artificial Intelligence (AI) has emerged as a rapidly evolving field with growing applications in medicine. AI was defined by Kurzweil in 1990 as "the art of creating machines that perform functions that require intelligence when performed by people" (9). In practical terms, AI systems are designed to analyze data, recognize patterns, and generate predictions or decisions that simulate aspects of human reasoning.

Artificial Intelligence encompasses several subfields, including machine learning (ML), deep learning (DL), natural language processing (NLP), and various neural network architectures (9,10). Machine learning is centered on the development of algorithms that can learn from data and progressively enhance their performance without being specifically programmed for each individual task (11,12). Depending on the structure of the data and the objective, ML approaches may be supervised, unsupervised, semi-supervised, or reinforcement-based (9,10).

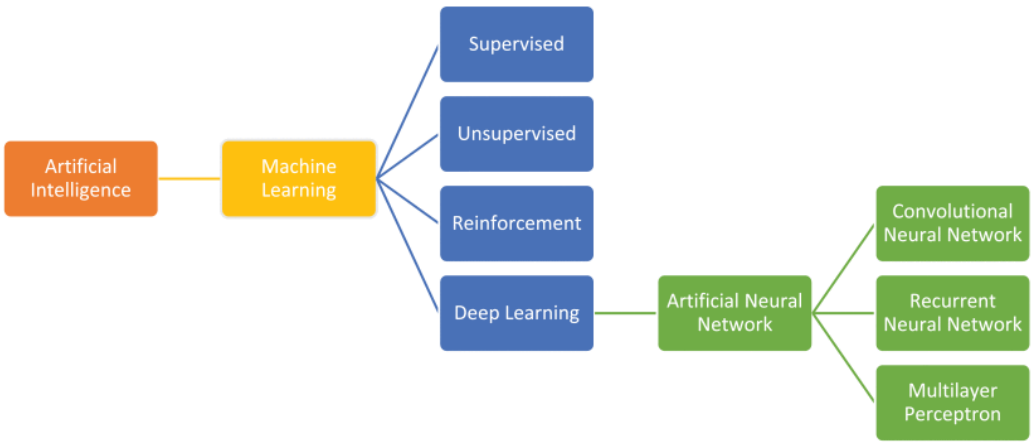


Figure 1. Artificial Intelligence and its different key subfields (9).

Deep learning, a subset of machine learning, is particularly relevant in healthcare applications. It relies on artificial neural networks composed of multiple interconnected layers that allow the automatic extraction of complex features from large datasets, especially images. Among these architectures, convolutional neural networks (CNNs) are widely used for medical image analysis, while other models such as recurrent neural networks (RNNs) are more commonly applied to sequential data (7,9).

In oncology, AI has been explored for various purposes, including early detection, image interpretation, risk stratification, outcome prediction, and decision support (7). In the field of oral cancer specifically, AI-based systems have been developed to analyze clinical photographs, radiographic images, histopathological slides, and even patient records through natural language processing. These tools aim to assist clinicians in distinguishing benign, potentially malignant, and malignant lesions, as well as in predicting prognosis and treatment response (2,7).

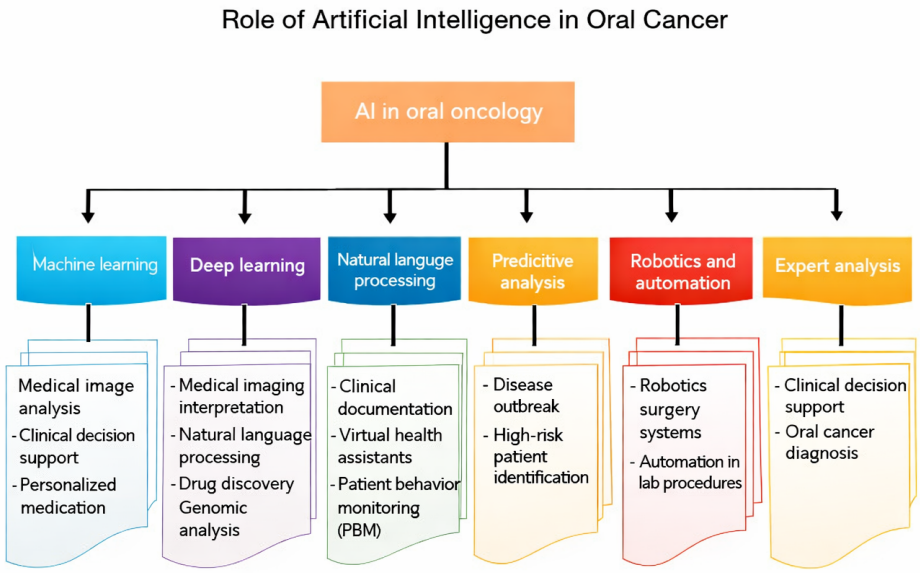


Figure 2. Role of Artificial Intelligence in Oral Cancer (adapted from Umapathy et al., 2024)(7).

However, although the number of studies investigating AI in oral oncology has increased in recent years, the real clinical usefulness of these technologies remains to be critically assessed. Many reported models demonstrate high accuracy values under controlled conditions,

yet questions persist regarding their generalizability, external validation, integration into routine dental practice, and ethical implications.

Therefore, the present study aims to examine whether Artificial Intelligence can be considered a useful tool in the detection, diagnosis, management and risk prediction of oral cancer. This literature review will analyze the current scientific evidence on AI-based systems applied to oral oncology, compare their diagnostic performance with conventional approaches when available, and discuss their potential advantages, limitations, and future perspectives in clinical dentistry.

2. OBJECTIVE

2.1 Primary Objective

The main objective of this documentary research is:

1. To assess the role of artificial intelligence in early detection, diagnosis, and management of oral cancer.

2.2 Secondary Objective

A secondary aim of this study is to evaluate how the diagnostic performance of AI-based methods compares to that of standard diagnosis technics.

The objectives of this research will be addressed through a systematic review conducted in accordance with PRISMA guidelines.

2.3 Research question

In patients affected by oral cancer, does the application of Artificial Intelligence improve diagnosis, treatment and clinical outcomes compared to standard methods?

2.4 Hypothesis

The application of artificial intelligence has the potential to improve diagnostic, therapeutic, and prognostic outcomes in oral cancer compared with existing conventional techniques.

2.5 Data measures

According to the PICO (P = population, I = intervention, C = comparison, O = outcome), the variables considered are defined as follows:

2.5.1 Population

The study population includes individuals diagnosed with oral cancer or presenting oral potentially malignant disorders such as leukoplakia, erythroplakia, oral submucous fibrosis, and oral lichen planus.

2.5.2 Intervention

Artificial Intelligence-based systems used for detection, diagnosis, classification, prognosis prediction, or treatment support. This includes machine learning, deep learning,

convolutional neural networks, artificial neural networks, and natural language processing models.

2.5.3 Comparison

The reference standard consisted of conventional diagnostic methods, including clinical examination, histopathological evaluation, and standard imaging interpretation by healthcare professionals.

In studies where a direct control group was not available, the performance of AI systems was compared to established diagnostic benchmarks reported within the same study.

2.5.4 Outcome

Primary outcomes included quantitative diagnostic performance metrics such as accuracy, sensitivity, specificity, AUC, and ROC analysis.

Secondary outcomes included clinical applicability, reproducibility, and potential impact on early detection and management.

3. MATERIAL AND METHODS

3.1 Eligibility Criteria

3.1.1 Inclusion Criteria

1. Publications exploiting AI or divisions of AI in the contexture of oral cancer or potentially malignant oral disorders were eligible.
2. Studies were eligible if they compared an AI system with a conventional diagnostic and or treatment approach. Research contrasting different AI models with each other was also included to broaden the evaluation of various AI techniques.
3. Any study involving participants that met the population criteria outlined in the PICO question was accepted.
4. Although randomised controlled trials were the preferred study design, the scarcity of research on AI-based diagnosis in oral cancer required the inclusion of high-quality observational studies, such as cohort and case-control designs, to better mirror real clinical settings.
5. Publications written in English were selected.
6. Publication from the last 5 years were selected.
7. For the IEEE Xplore database, only journal articles were included.
8. Studies reporting at least one quantitative performance metric (accuracy, sensitivity, specificity, AUC, ROC curve, F1-score, etc.).

3.1.2 Exclusion Criteria

1. Studies that did not pertain to oral cancer or potentially malignant oral disorders were omitted.
2. Studies lacking a focus on artificial intelligence was also excluded.
3. Any study failing to align with the criteria established in the PICO research question was removed.
4. Studies were excluded if they did not report at least one quantitative performance indicator, including accuracy, sensitivity, specificity, area under the curve (AUC), receiver-operating characteristics curve (ROC), needed to evaluate diagnostic effectiveness.
5. Secondary evidence, including literature reviews and meta-analysis was excluded to ensure the analysis relied solely on primary evidence and to prevent duplication of study finding.

3.2 Information sources

These two electronic databases were used:

1. Medline complete
2. IEEE Xplore

These two databases were chosen. Medline complete is a leading biomedical database of references for medicine, health sciences, dentistry, and related fields. It provides essential scientific literature for evidence-based research. IEEE Xplore is a leading digital library that provides access to a wide range of technological and engineering research publications. In this review, non-academic sources were excluded in order to ensure the reliability and reproducibility of the study.

3.3 Search Strategy

A structured search strategy was developed by combining appropriate keywords with Boolean operators, following the PICO framework.

The corresponding search equation for the Medline Complete database is presented below: (“artificial intelligence or ai or a.i or machine learning or deep learning or generative ai”) AND (“oral cancer or carcinoma in oral cavity or malignancy in oral cavity or mouth cancer”) AND (“detection or diagnosis or identification or early detection or recognition”) AND (“management or treatment or intervention or therapy”) AND (“prognosis or outcome or recovery or predictor or survival”) NOT (meta-analysis or systematic review or literature review or meta-analysis or overview or review or meta-synthesis or meta synthesis”). A total of 76 results were obtained.

The IEEE Xplore search question was formulated as follows: (“artificial intelligence”) AND (“oral cancer”) NOT (“meta-analysis”) NOT (“systematic review”). A total of 144 results were obtained. The filter used was for journals, resulting in 29 studies.

3.4 Study Selection process

Titles and abstracts were assessed to select potentially relevant studies.

To perform the screening, the full texts of the selected articles were manually obtained and evaluated according to the eligibility criteria.

An overview of the study selection process is provided in the PRISMA flow diagram shown in Figure 3.

3.5 Data collection process

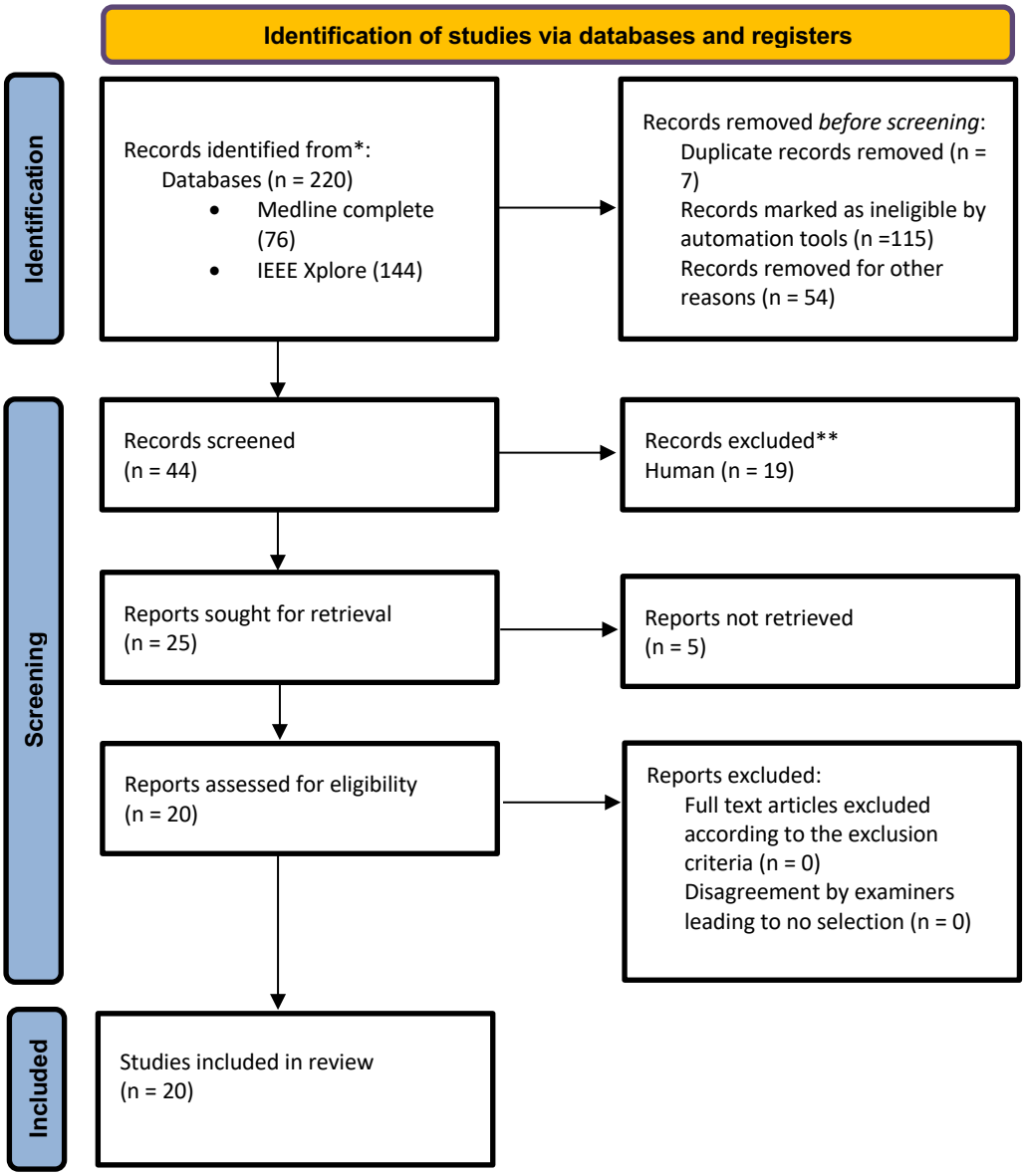


Figure 3. PRISMA flow diagram illustrating the different stages of the study selection process.

4. RESULTS

4.1 Presentation of results in a tabular form

Table 1. Summary of the included studies, presenting the authors, year of publication, study design, and the different AI-related tasks addressed.

	Author and year of publication	Type of study	AI task
1	*Prabhune et al. 2025 (13)*	Experimental study	Classification
2	*Jung et al. 2026 (14)*	Retrospective cohort study	Prognosis/Risk prediction
3	*Monani et al. 2025 (15)*	Methodological study	Prognosis/Risk prediction
4	*Coole et al., 2023 (16)*	Clinical trial	Diagnosis
5	*Adeoye and Su. 2026 (17)*	Retrospective cohort study	Prognosis/Risk prediction
6	*Huang et al. 2025 (18)*	Retrospective cohort study	Prognosis/Risk prediction
7	*Tanriver et al. 2021 (19)*	Cross-sectional study	Screening and classification
8	*Flügge et al. 2023 (20)*	Cross-sectional study	Binary detection
9	*Zhang et al., 2025 (21)*	Experimental study	Segmentation and classification
10	*Lan et al., 2024 (22)*	Retrospective study	Diagnosis and prognosis
11	*Aubreville et al., 2017 (23)*	Experimental study	Diagnosis and classification
12	*Akram et al., 2026 (24)*	Experimental study	Classification
13	*Shephard et al., 2024 (25)*	Retrospective study	Prognosis/Risk prediction
14	*Li et al., 2025 (26)*	Observational study	Classification/Risk prediction
15	*Shephard et al., 2025 (27)*	Retrospective study	Diagnosis and prognosis
16	*Xu et al., 2019 (28)*	Experimental study	Diagnosis and classification
17	*Goswami et al., 2024 (29)*	Computational study	Classification
18	*Devindi et al., 2024 (30)*	Experimental study	Detection and classification
19	*Ul Haq et al. 2023 (31)*	Experimental study	Classification and Diagnosis
20	*Ramya et al. 2025 (32)*	Experimental study	Classification and diagnosis

Table 2. The table outlines the main characteristics of the selected studies, including authors, year of publication, study design, and the AI applications examined.

Authors and year	AI model	Clinical application	Data type	Reported performance metrics	Reference method	Principal findings	Study limitations
Prabhune et al. 2025 (13)	U-Net + DenseNet201	To segment lesions and classify (normal inflammatory, premalignant or carcinoma)	794 intraoral photos	Accuracy 94%; Sensitivity: 0,83 Specificity: 0,71 F1-score: 0,67	Expert-labelled clinical image	Good performance of mobile application for AI-assisted OPMDs screening by non-specialist health workers	Small dataset; moderate specificity; usability and workflow still being evaluated
Jung et al. 2026 (14)	XGBoost machine learning model (RADAR score)	Predict recurrence and survival after surgery in head and neck cancer	Clinical data + laboratory biomarkers	Overall survival prediction AUC: 0.788-0.820; Recurrence-free survival prediction AUC: 0.785-0.831 Sensitivity: 72-81% Specificity: 68-76%	Cross-validation within multicenter dataset	AI model accurately predicted recurrence and survival up to 5 years using dynamic surveillance data	Retrospective dataset; no radiomics, genomic data; prospective validation required
Monani et al. 2025 (15)	CNN with transfer learning + SMOTE	Detect oral cancer (binary classification)	Clinical photographic images	Accuracy: 81.38% Precision: 84.62% Sensitivity: 78.57% F1-score: 81.48% AUC: 0.908	Compared with baseline CNN performance reported in medical imaging literature	Good performance of CNN model for early oral cancer detection using transfer learning	Small dataset; no external validation; possible overfitting

Authors and years	AI model	Clinical application	Data type	Reported performance metrics	Reference methods	Principal findings	Study limitations
Coole et al. 2023 (16)	Multitask deep learning network (MTN) integrated in multimodal imaging system	Detect high-risk oral lesions and guide biopsy location	Autofluorescence wide-field imaging + micro-endoscopy images	Imaging resolution: 110micron projection resolution: 400micron probe tracking error <1.5 mm	Histopathological analysis of biopsies	AI system enables real-time projection of cancer risk and biopsy guidance during oral examination	Pilot clinical study; small sample size; needs larger validation
Adeoye and Su. 2026 (17)	Multitask AI model (TabPFN, TabNet, XGBoost, Random Forest)	Predict treatment failure, malignant progression and recurrence in OPMDs after surgery	Clinical and histopathological data (tabular dataset)	Treatment failure AUC: 0.829 Malignant progression AUC: 0.912 Recurrence AUC: 0.791	WHO dysplasia grading and binary dysplasia grading	AI model outperformed traditional dysplasia grading systems for risk prediction	Retrospective cohort; external validation; dataset relatively small; further multicenter validation needed
Huang et al. 2025 (18)	XGBoost machine learning model	Predict recurrence after adjuvant treatment in oral cavity cancer	Clinical and treatment data (tabular dataset)	Accuracy: 57,9% Sensitivity: 57,1% AUC: 0,58	Cox regression analysis	Pathological lymph node status was the most important predictor of recurrence	Small dataset (187 patients); limited testing set; model performance relatively low

Authors and years	AI model	Clinical application	Data type	Reported performance metrics	Reference methods	Principal findings	Study limitations
Tanriver et al. 2021 (19)	Deep learning CNN models (U-Net, Mask R-CNN, YOLOv5, EfficientNet-b4, Inception-v4)	Detection and classification of oral lesions (benign, OPMDs, carcinoma)	Photographic images of oral cavity	Segmentation F1: 0,929 Detection AP: 0,644 Classification F1:0.858	Histopathology-labelled dataset and expert annotations	EfficientNet-b4 and Inception-v4 achieved best classification performance; YOLOv5 effective for real-time lesion detection	Small dataset; high image variability (lighting, angle, quality)
Flügge et al. 2023 (20)	Vision Transformer (Swin-Trans former)	Detect OSCC from clinical images	Clinical photographs of oral cavity	Accuracy: 0.986 AUC: 0.99 Sensitivity: 0.986 Specificity: 0.986 F1-score: 0.986	Histopathologicaly confirmed diagnosis used as ground truth	AI model showed very high performance for OSCC detection on clinical photograph	Single-center dataset; high-quality images; real-world performance unknown
Zhang et al. 2025 (21)	gamUnet (U-Net + Global Attention Mechanism), gamResNet	OSCC segmentation and classification	Histopathological images	Segmentation F1: 0.909 Accuracy: 0.9184. Classification (x100): AUC: 0.991, Accuracy:0.902 F1: 0.45 Classification (x400): AUC: 0.924 Accuracy: 0.899	Compared with U-Net, Attention-U-Net, Res-U-Net, MobilNet, EfficientNet, ResNet	Global attention improved OSCC segmentation and classification performance vs standard CNN model	Dataset size limited; mainly experimental dataset; real clinical validation needed

Authors and year	AI model	Clinical application	Data type	Reported performance metrics	Reference methods	Principal findings	Study limitations
Lan et al. 2024 (22)	ResNet50 + radiomics features	Predict occult cervical lymph node metastasis (OCLNM) and prognosis	MRI (T1-weighted and T2-weighted images)	Training: AUC: 0.928 Accuracy: 0.861. Test: AUC: 0.878 Accuracy: 0.868. External validation: AUC: 0.796 and 0.834	Compared with radiomics-only models and other CNN models (ResNet18, ResNet34, EfficientNet, DenseNet)	ResNet50: radiomics achieved best performance for predicting occult lymph node metastasis	Retrospective study; manual tumor segmentation; need prospective validation
Aubreville et al. 2017 (23)	Convolutional Neural Network (CNN) + patch probability fusion; Inception v3 (transfer learning)	Automatic classification of cancerous vs normal oral tissue	Confocal Laser Endimicroscopy (CLE) images	Accuracy: 88.3% sensitivity: 86.6% specificity: 90.0% AUC: 0.955	Compared with texture-based ML (LBP+Random Forest, GLCM features)	CNN significantly outperformed traditional texture-based classifiers for OSCC detection	Small dataset (12 patients); artifact-affected images; lack of healthy control group
Akram et al. 2026 (24)	Deep CNN model (EfficientNetB3, DenseNet121, ResNet50)-Deep Visual Detection System (DVDS)	Automatic detection of OSCC and leukoplakia	Histopathological images	EfficientNetB3: Accuracy: 97.05% F1score: 97.05% Specificity: 97.17% Sensitivity: 96.92%	Compared between EfficientNetB3, DenseNet121 and ResNet50	EfficientNetB3 achieved the highest performance for OSCC detection and multi-class lesion classification	Dataset derived from public databases; lack of external clinical validation; potential dataset imbalance

Authors and year	AI model	Clinical application	Data type	Reported performance metrics	Reference methods	Principal findings	Study limitations
Shephard et al. 2024 (25)	Deep learning pipeline (HoVer-Net+ segmentation + MLP classifier)	Predict malignant transformation risk of oral epithelial dysplasia (OED)	Whole slide histopathological images	Internal validation: AUROC: 0.77 Recall: 0.84 F1-score: 0.57 External validation: AUROC: 0.75 Recall: 0.92 F1-score: 0.69	Compared with pathologist grading systems (WHO grade and Binary grade)	AI model predicted malignant transformation risk and showed comparable or better sensitivity than traditional histopathological grading	Relatively small dataset; retrospective study; single-center training cohort; segmentation generalization issues on external datasets
Li et al. 2025 (26)	OMMT-PredNet (ResNet50 + CBAM multimodal deep learning)	Identification of oral epithelial dysplasia and prediction of malignant transformation in oral leukoplakia	High-resolution clinical photographs + clinical records	MT prediction: AUC: 0.9592 Sensitivity: 0.8549 Specificity: 0.9490 F1 0.8972. OED detection: AUC: 0.9219 Sensitivity: 0.7931 Specificity: 0.9182 F1: 0.8550. External validation: AUC: 0.9469	Internal test cohort + external validation cohort	Multimodal AI accurately predicts malignant transformation risk and identifies OED from clinical images and clinical data	Retrospective study, limited dataset, need for larger multicenter validation

Authors and year	AI model	Clinical application	Data type	Reported performance metrics	Reference methods	Principal findings	Study limitations
Shephard et al. 2025 (27)	Transformer-based pipeline (ODYN: Trans-UNet + Hover-Net+ + MLP)	Detect oral epithelial dysplasia and predict malignant transformation	Whole slide histopathology images	Segmentation: F1: 0.81 (internal) F1: 0.71 (external) OED classification: F1: 0.96 AUROC: 0.93 Transformation prediction: AUROC: 0.71 (internal) AUROC: 0.73 (external)	Compared with clinical grading systems (WHO grading and binary grading)	AI achieved performance comparable to clinical grading systems for predicting malignant transformation	Retrospective dataset, domain shift between cohorts, limited external controls, need prospective validation
Xu et al. 2019 (28)	3D Convolutional Neural Network (3DCNN) and 2D CNN	Early diagnosis of oral cancer (benign vs malignant classification)	CT images	3DCNN: AUC: 0.796 Accuracy: 0.754 Sensitivity: 0.818 Specificity: 0.739. 2DCNN: AUC: 0.734 Accuracy: 0.687 Sensitivity: 0.707 Specificity: 0.632	Compared with a 2D CNN model	3DCNN improved diagnostic performance compared with 2DCNN by using 3D spacial information from CT images	Small dataset (7000CT images), retrospective data, only CT modality evaluated

Authors and year	AI model	Clinical application	Data type	Reported performance metrics	Reference methods	Principal findings	Study limitations
Goswami et al. 2024 (29)	Light Gradient Boosting Machine (LightGBM)	Classification of oral cancer and precancerous lesions	White-light oral photographs	Binary classification: Accuracy: 99.25% Precision: 99.18% Recall: 99.31% F1: 99.24% Specificity: 99.31%. Multiclass: Accuracy: 98.88% Precision: 98.86% Recall: 97.92% F1: 98.38% Specificity: 99.03%	Compared with NB, SVM, Random Forest, and MLP classifiers	LightGBM achieved the best performance. Combining color + texture features produce	Small datasets, images partly collected from internet, limited external validation
Devindi et al. 2024 (30)	Multimodal Deep CNN (MobileNetV3-Large, DenseNet121, ResNet50, MixNet-S, HRNet, Inception-V3)	Early detection of oral cancer / OPMDs classification	Smartphone oral cavity images	Accuracy: 81% Precision: 79% Recall: 79% F1-score: 78	Compared with single-modal models (image only) and late-fusion model	Multimodal model improved classification compared to image-only approaches	Small dataset, single population, limited external validation

Authors and year	AI model	Clinical application	Data type	Reported performance metrics	Reference methods	Principal findings	Study limitations
Ul Haq et al. 2023 (31)	Hybrid AI: Gabor filter + ResNet50 + CatBoost	Classification of OSCC	Histopathological biopsy images	Accuracy: 94.92% Precision: 95.51% Sensitivity: 94.30% Specificity: 95.54% F1-score: 94.90% AUC: 94.90%	Compared three approaches: Gabor+CatBoost, ResNet50+CatBoost, Hybrid mode	Hybrid model achieved highest performance for OSCC detection	Single dataset, no external validation, limited clinical testing
Ramya et al. 2025 (32)	XSFN-Net (Xception + Deep Spiking Neural Network + Fractional Calculus)	Classification of oral squamous cell carcinoma (OSCC)	Histopathological images	Accuracy: 97.54% Sensitivity (TPR): 96.60% Specificity (TNR): 95.54%	Compared with CNN, OralNet, MTGO-DL, DCNN	Proposed model achieved highest classification performance among compared model	Small dataset, limited real clinical validation

4.2 Characteristics of the included studies

A total of 20 primary studies met the eligibility criteria and were included in this review. The studies were published between 2017 and 2026 and comprised experimental studies, retrospective cohort studies, cross-sectional studies, observational studies, computational studies, and one clinical trial.

Several studies focused on classification of lesions (for example, normal versus inflammatory versus premalignant versus carcinoma, or benign versus malignant). Others addressed screening and binary detection, segmentation, diagnosis, and risk prediction or prognosis, while a number of works combined more than one task, such as diagnosis with classification or diagnosis with prognosis. The distribution of studies according to the main AI task is shown in figure 4.

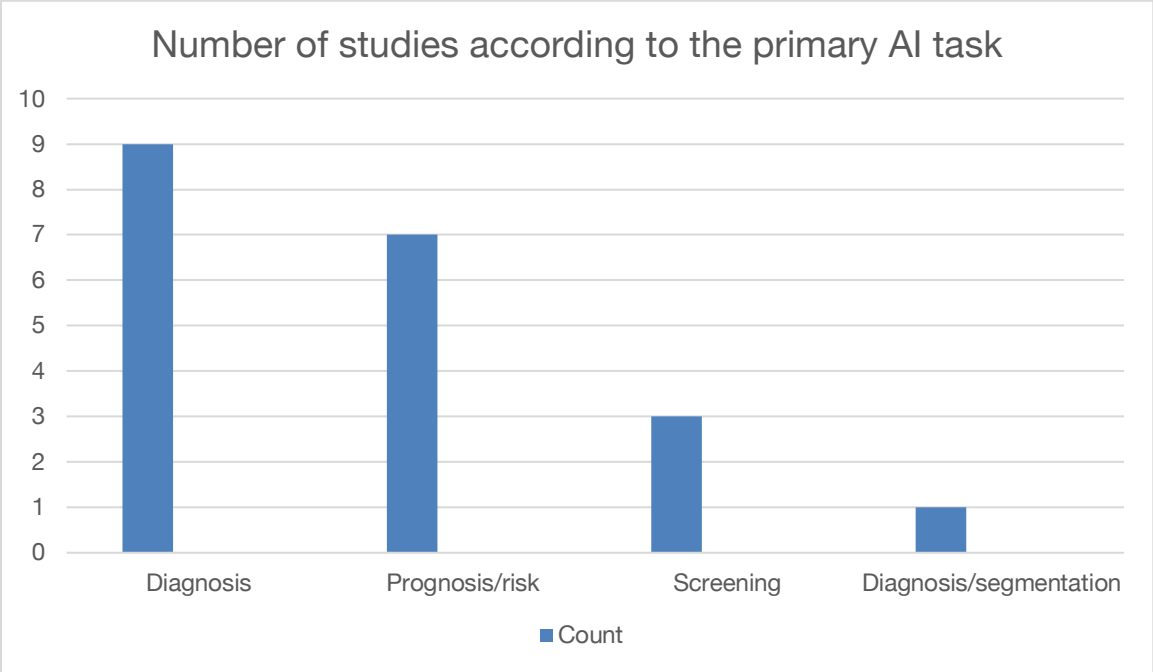


Figure 4. Number of included studies according to the primary AI task.

4.3 Artificial Intelligence models and input data

The studies used a wide range of AI models. Deep learning architectures included convolutional neural networks (CNN), U-Net-based models, DenseNet201, and multitask deep learning networks. Other approaches involved machine learning methods such as XGBoost, Random Forest, TabNet, and TabPFN, as well as different hybrid or ensemble models applied to tabular clinical data.

There was also variability in the input data used to train and test these models. Several studies employed clinical intraoral photographs or images acquired with mobile devices for screening, detection, and classification of oral lesions. Others used histopathological images or whole-slide images to assist in diagnosis and grading. Additional works relied on radiological images (such as computer tomography or other imaging modalities) or on tabular datasets containing clinical, histopathological, laboratory, or treatment variables, sometimes in a multimodal format. These data modalities are shown in figure 5.

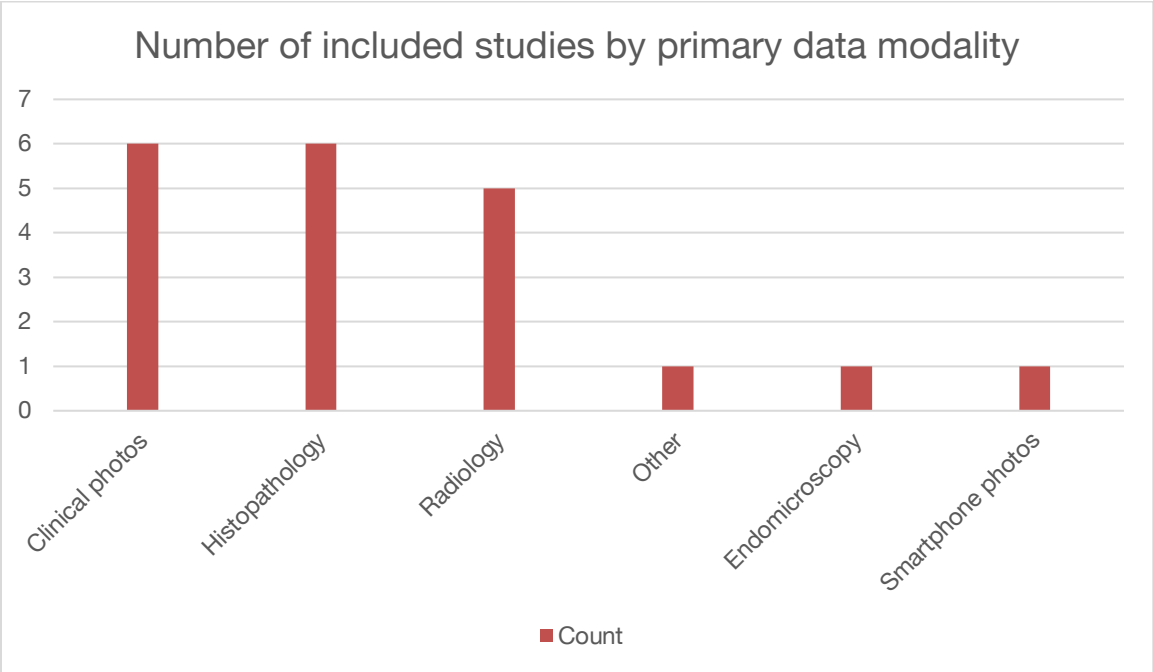


Figure 5. Number of included studies by primary data modality.

4.4 Reported diagnostic and prognostic performance

All included studies reported at least one quantitative performance metric such as accuracy, sensitivity, specificity, precision, F1-score, AUC, or ROC analysis. In image-based classification or detection tasks using clinical photographs, reported accuracies frequently exceeded 80%, and several models achieved AUC values close to or above 0.90. Sensitivity and specificity values varied between studies but were generally presented alongside accuracy and F1-score to describe model performance.

Histopathology-based models reported metrics including accuracy, sensitivity, specificity, and AUROC for distinguishing cancerous from non-cancerous tissues or for classifying different lesion categories. For prognostic and risk-prediction tasks using clinical and histopathological variables, AUC values were commonly reported for outcomes such as treatment failure, malignant progression, recurrence, overall survival, or recurrence-free survival. In some of these studies, the AI models were

directly compared with reference standards such as WHO dysplasia grading systems or Cox regression analysis.

4.5 Reference standards and comparisons

Most studies used expert assessment as the reference standard. For image-based tasks, ground truth labels were typically provided by specialists through clinical examination or histopathological confirmation. In prognostic studies, outcomes such as recurrence or treatment failure were defined based on clinical records and follow-up data.

Some studies included comparisons between AI model performance and established grading systems or statistical models, while others reported only internal validation results without direct comparison to clinicians.

4.6 Reported limitations within the studies

Several studies explicitly described limitations related to their design or data. Commonly mentioned issues were small sample size, single-center datasets, lack of external validation, and restricted image acquisition conditions.

In some cases, authors highlighted moderate performance in certain metrics, the need for larger or more diverse datasets, and the absence of prospective evaluation or real-world tes

5. DISCUSSION

5.1 AI for screening and early detection

Several studies in this review applied AI to intraoral or smartphone photographs to screen for normal mucosa, inflammatory lesions, OPMD and OSCC. Prabhune *et al.* developed a U-Net+DenseNet201 model embedded in a mobile application to segment lesions and classify them as normal, inflammatory, premalignant or carcinoma. They achieved high accuracy on clinical images from a tertiary center and community camps (13). Tanriver *et al.* evaluated deep learning models including U-Net, Mask R-CNN, YOLOv5, EfficientNet-B4 and Inception-V4 for detection, segmentation and classification of oral lesions, reporting good performance for both screening and classification tasks (19). Other works, such as those by Flügge *et al.* Goswami *et al.* and Devindi *et al.* assessed CNNs, vision transformers and multimodal deep networks for OSCC and OPMD detection using white-light or smartphone-based images(20,29,30).

These screening systems suggest that AI-based screening tools are promising as complementary aid for non-specialist dentists or community health workers, especially in settings with limited access to oral medicine experts (13,19,20,29,30). They proved useful during routine examinations by highlighting and distinguishing suspicious lesions and supporting referral decisions.

However, these results should be interpreted with caution. Most models were trained on relatively small, single-center datasets, retrospectively collected and with highly curated, high-quality images and limited variability in acquisition conditions, which may overestimate performance compared with everyday examinations in primary care (13,19,20,29,30). External validation on independent cohorts and devices was uncommon, and the performance of these systems in heterogeneous real-world clinical environments remains uncertain.

Thus, before these systems can be recommended for widespread implementation, larger prospective studies in diverse populations, using standardized image acquisition protocols and direct comparison with general dentists, are required to determine whether AI-assisted screening actually increases early detection rates or reduces unnecessary referrals.

5.2 AI for diagnosis, segmentation and classification

Beyond initial screening, several studies focused on diagnostic and segmentation tasks using clinical, endomicroscopic or histopathological images. Coole *et al.* integrated a multitask deep learning network into an autofluorescence and micro-endoscopy imaging system to identify high-risk oral lesions and guide biopsy location in real time during oral examination (16).

Aubreville *et al.* designed a CNN-based approach using patch-probability fusion and an Inception-V3 architecture to automatically distinguish cancerous from normal oral tissue on confocal laser endomicroscopy images, outperforming traditional feature-based classifiers (23).

Histopathology oriented work by Zhang *et al*., Akram *et al*., Ul Haq *et al.* and Ramya *et al.* employed architectures such as U-Net variants, ResNet, EfficientNet and hybrid deep networks for OSCC segmentation and multi-class classification on biopsy or whole-slide images, generally reporting high values of accuracy, sensitivity, specificity and AUROC (21,24,31,32).

These diagnostic models indicate that AI can capture subtle morphological patterns and may help reduce inter-observer variability in histological assessment. Nevertheless, most of these studies were experimental, relied on restricted datasets from one institution and rarely included robust external validation, which limits their immediate applicability in routine pathology laboratories.

Radiological diagnosis was addressed by Xu *et al*., who used 2D and 3D CNNs on dynamic contrast-enhanced CT images to classify benign versus malignant oral lesions. Lan *et al.* combined a ResNet50 backbone with radiomics features extracted from MRI to predict occult cervical lymph node metastasis in oral cancer (22,28).

Both studies reported improved diagnostic performance compared with simpler models or conventional parameters, but they also shared the limitations of retrospective, single-center design and limited sample sizes.

Overall, diagnostic and segmentation models show that AI can identify subtle morphological and radiological patterns and may help standardize interpretation across observers. In particular, histopathology and advanced imaging applications suggest that AI could support pathologists and radiologists in distinguishing benign from malignant lesions and in delineating tumor extent (16,21–24,28,31,32).

Nevertheless, most available studies are experimental or retrospective, rely on restricted institutional datasets and have not yet evaluated the real impact of AI on clinical decision making or diagnostic consistency between clinicians.

The lack of standardization is another challenge. Differences in model design, training strategies, and performance reporting make it difficult to reproduce results and establish clear benchmarks across studies. Their current role should be considered supportive rather than as a replacement for conventional expert assessment and biopsy (6–9,16,21–24,28,31,32).

5.3 AI for prognosis and risk prediction

A relevant group of studies focused on prognostic tasks, including prediction of recurrence, malignant transformation, treatment failure and survival outcomes. Jung *et al.* developed an XGBoost-based RADAR score using clinical and laboratory biomarker data to estimate overall and recurrence-free survival after head and neck cancer surgery, achieving good AUC values in internal validation (14).

Adeoye and Su constructed a multitask model that combined TabPFN, TabNet, XGBoost and Random Forest algorithms to predict treatment failure, malignant progression and recurrence in surgically treated OPMDs. In their analysis, the AI model outperformed WHO dysplasia grading and binary dysplasia grading for risk prediction (17).

Huang *et al.* applied an XGBoost model to clinical and treatment variables in oral cavity cancer to predict recurrence after adjuvant therapy, identifying pathological lymph node status as the most important predictor, although the overall accuracy and AUC were only moderate (18).

Shephard *et al.* proposed deep learning pipelines combining HoVer-Net with multilayer perceptrons and a transformer-based ODYN framework on whole-slide histology to detect oral epithelial dysplasia and to estimate malignant transformation risk, reporting AUROC values that were comparable to or higher than conventional grading systems in both internal and external cohorts (25,27).

Li *et al.* introduced a multimodal network (OMMT-PredNet) that integrates high-resolution clinical photographs with clinical data using a ResNet50 backbone and attention mechanisms to detect OED and predict malignant transformation in oral leukoplakia (26). Their model obtained high AUC, sensitivity and specificity in internal and external validation sets, suggesting that multimodal AI could support personalized risk stratification and tailored surveillance of patients with OPMD.

From a clinical perspective, these prognostic and risk-prediction models illustrate the potential of AI to refine risk stratification beyond traditional clinical and histopathological factors and to guide more individualized follow-up strategies for the patient with OPMDs and oral cancer.

However, despite these promising findings, all prognostic models in this review were retrospective, mostly single-center and based on limited sample sizes and follow-up durations. None reported prospective evaluation of their impact on clinical decision-making or patient outcomes, it remains unknown whether using these models would actually lead to earlier detection of recurrence, more appropriate surveillance intervals or improved survival (14,17,18,25–27).

At present, AI-based risk scores should therefore be viewed as exploratory tools that may complement, but not replace, established prognostic systems and multidisciplinary clinical judgement.

5.4 Limitations of this review

This systematic review has several limitations that should be acknowledged. The search was limited to two databases (Medline Complete and IEEE Xplore), to studies published in English and to a five-year period, which may have led to the exclusion of relevant articles published in other languages, databases or time frames.

For IEEE Xplore, only journal articles were included, so conference proceedings important sources in AI development were not considered.

In this review, a specific tool designed to systematically evaluate the risk of bias in AI studies was not applied, and it is likely that studies reporting poor or negative results are under-represented in the published literature. Consequently, the reported performance of AI systems may appear higher in the current evidence base than would be expected in everyday clinical practice.

Beyond methodological issues, there are also important conceptual limitations that affect the practical usefulness of current AI systems for oral cancer. Most published models provide limited information about how predictions are generated, which reduces explainability and can hinder clinician understanding and acceptance (7–9).

The lack of standardized protocols for image acquisition, data preprocessing, model training and reporting makes it difficult to compare results across studies and to reproduce performance in different clinical environments.

Many datasets are derived from single centers or specific populations, raising concerns about hidden biases and the risk that models may perform less well in under-represented groups, thereby potentially amplifying existing disparities in access and quality of care (2,3,6–9,12).

Finally, insufficient attention has been paid to ethical, legal and data-protection aspects, such as privacy, informed consent, accountability and regulatory oversight, which further limits trust in these tools and complicates their safe integration into routine oral healthcare (7–9,12).

5.5 Implications for future research

Future research on AI in oral oncology should respond to the high global burden of oral cancer and the need for earlier detection and more consistent diagnosis (1–8). Screening models based on intraoral or smartphone images must be trained and validated on large, multi-center datasets that include different populations, risk profiles and real-world image conditions, using standardized protocols for image capture and expert labelling. This will help determine whether AI can truly detect subtle lesions earlier in primary care and community settings, not only in tertiary centers (1–3,5–8,13,19,20,29,30)

For diagnostic and segmentation tasks in histopathology and radiology, future studies should use harmonized multi-institutional datasets to test whether promising models maintain their performance across laboratories, scanners and imaging protocols (2,3,6,7,12,16,21–24,28,31,32).

Prospective clinical studies are needed to compare AI-assisted diagnosis with standard practice and to measure outcomes such as agreement between observers, biopsy selection, diagnostic delay and impact on treatment planning. Prognostic research should move from retrospective model development to prospective evaluation of whether AI-based risk scores actually change follow-up strategies and improve patient outcomes in oral cancer and OPMD (2,3,6–8,14,17,18,25–27).

Across all applications, future work should follow transparent and standardized reporting for data sources, preprocessing, model design and evaluation metrics, so that results can be reproduced and different systems can be fairly compared (2,3,6,7,9–12,16,21–24,28,31,32).

Studies design, involving dentists, oral surgeons, pathologists, oncologists and patients, will be important to ensure that AI tools are not only accurate but also acceptable, easy to integrate into existing workflows and trustworthy (4,6–9,12).

Finally, implementation and cost-effectiveness studies will be necessary to determine how AI can be used as a complementary tool alongside conventional diagnostic methods to reduce diagnostic delays and ultimately improve survival and quality of life in patients with oral cancer and OPMDs.

6. CONCLUSIONS

The findings of this systematic review indicate that Artificial Intelligence has significant potential as a supportive tool in the detection, diagnosis and management of oral cancer and oral potentially malignant disorders.

Across the 20 included studies, AI-based systems demonstrated encouraging performance for a wide spectrum of tasks, including screening, lesion segmentation, histopathological and radiological interpretation, and prognostic risk estimation.

Image-based models applied to clinical and smartphone photographs frequently achieved high accuracy and AUC values for distinguishing normal mucosa, inflammatory lesions, OPMDs and oral squamous cell carcinoma, suggesting a possible role in supporting non-specialist clinicians and community health workers.

Histopathology- and radiology-focused approaches also showed that AI can capture subtle morphological and imaging patterns, sometimes matching or surpassing conventional grading systems and statistical models in predicting malignant transformation, recurrence and survival.

However, the current evidence base presents important limitations that restrict the immediate clinical implementation of these tools. Most studies were retrospective, based on single-center datasets with limited sample sizes, and lacked rigorous external validation or prospective assessment of their impact on patient outcomes.

Heterogeneity in image acquisition, data preprocessing, model development and reporting further complicates the comparison between systems and raises concerns about generalizability to diverse clinical environments.

In addition to methodological constraints, there are conceptual and ethical challenges related to transparency, explainability, data protection and potential biases embedded in training datasets, which may affect clinician trust and could inadvertently exacerbate existing inequalities in access to care.

For these reasons, AI should currently be viewed as an adjunct to, rather than a replacement to standard methods as clinical examination, histopathological confirmation and multidisciplinary decision-making in oral oncology.

Future research should prioritize large, multi-centers and prospective studies that validate AI models in real-world settings and assess their influence on diagnostic accuracy, timing of diagnosis, treatment planning and patient prognosis.

Standardized frameworks for data collection, model evaluation and reporting, together with robust ethical and regulatory guidelines, will be essential to integrate AI safely and effectively into dental practice. If these challenges are addressed, AI has the potential to contribute meaningfully to earlier detection, more consistent diagnosis and more personalized follow-up of patients with oral cancer and OPMDs.

7. SUSTAINABILITY

The incorporation of Artificial Intelligence into oral cancer care raises not only clinical and ethical questions but also considerations related to sustainability in healthcare systems.

From an environmental perspective, AI models require computational resources for training, updating and deployment, which are associated with energy consumption and a corresponding carbon footprint, especially in the case of deep learning architectures trained on large imaging datasets. Strategies such as model optimization, the use of more efficient algorithms, and the adoption of energy-efficient hardware and cloud infrastructures can help mitigate this impact.

At the same time, if appropriately validated and implemented, AI could contribute to a more sustainable use of healthcare resources by improving early detection, reducing unnecessary referrals or biopsies and optimizing follow-up schedules based on individual risk profiles.

Earlier diagnosis and better risk stratification may help to decrease the need for aggressive treatments and long hospital stays, with potential benefits both for patients and for the overall efficiency of health services.

In dentistry, AI-assisted screening using smartphones or simple imaging devices could support outreach programs in underserved areas, favoring more equitable access to early diagnosis.

Social and organizational sustainability also requires that AI tools be acceptable, usable and trustworthy for dentists, oral surgeons, oncologists and patients.

This implies not only technical robustness but also clear communication of model capabilities and limitations, adequate training for clinicians, and mechanisms to maintain human oversight and accountability. Without these elements, there is a risk that AI systems may increase workload, generate mistrust or widen disparities between centers that can afford advanced technologies and those that cannot.

Finally, sustainable integration of AI in oral oncology must respect data protection regulations and ethical principles, ensuring that patient data are collected, stored and used securely and transparently.

Collaborative multicenter initiatives that share anonymized data under robust governance frameworks can reduce duplication of efforts, improve model generalizability and avoid the proliferation of many small, non-reusable datasets.

In this way, AI research and implementation in oral cancer can evolve towards a model that is not only technologically advanced, but also environmentally responsible, socially just and economically viable for healthcare systems.

8. REFERENCES

1. Chen R, Wang Q, Huang X. Intelligent deep learning supports biomedical image detection and classification of oral cancer. Technol Health Care. 2024;32:465–75.

2. Khanagar SB, Alkadi L, Alghilan MA, Kalagi S, Awawdeh M, Bijai LK, et al. Application and Performance of Artificial Intelligence (AI) in Oral Cancer Diagnosis and Prediction Using Histopathological Images: A Systematic Review. Biomedicines. 2023;11(6):1612.

3. Vinay V, Jodalli P, Chavan MS, Buddhikot ChaitanyaS, Luke AM, Ingafou MSH, et al. Artificial Intelligence in Oral Cancer: A Comprehensive Scoping Review of Diagnostic and Prognostic Applications. Diagnostics. 2025;15(3):280.

4. Adeoye J, Thomson P. A call for an established oral cancer classification by etiology and revision of related terminology. Oral Dis. 2022;28(3):840–2.

5. Kouketsu A, Doi C, Tanaka H, Araki T, Nakayama R, Toyooka T, et al. Detection of oral cancer and oral potentially malignant disorders using artificial intelligence-based image analysis. Head & Neck. 2024;46(9):2253–60.

6. Mirfendereski P, Li GY, Pearson AT, Kerr AR. Artificial intelligence and the diagnosis of oral cavity cancer and oral potentially malignant disorders from clinical photographs: a narrative review. Front Oral Health. 2025;6:1569567.

7. Umapathy VR, Natarajan PM, Swamikannu B, Jaganathan S, Rajinikanth S, Periyasamy V. Role of Artificial Intelligence in Oral Cancer. Wu Y, editor. Adv Public Health. 2024;2024(1):3664408.

8. Kharche A, Mathur A, Mehta V. AI-powered oral cancer detection: A breakthrough in dental diagnostics. Oral Oncol Rep. 2024;10:100293.

9. Nazar M, Alam MM, Yafi E, Su'ud MM. A Systematic Review of Human–Computer Interaction and Explainable Artificial Intelligence in Healthcare With Artificial Intelligence Techniques. IEEE Access. 2021;9:153316–48.

10. Nath S, Raveendran R, Perumbure S. Artificial Intelligence and Its Application in the Early Detection of Oral Cancers. Clin Cancer Investig J. 2022;11(1):5–9.

11. Mana SC, Kalaiarasi G, R Y, Helen LS, Senthamil Selvi R. Application of Machine Learning in Healthcare: An Analysis. In: Proc 3rd Int Conf Electron Sustain Commun Syst (ICESC). 2022[Internet]. Coimbatore, India: IEEE; 2022 [cited 2025 Dec 10]. p. 1611–5.

12. Salwaji S, Pasupuleti MK, Manyam R, Pasupuleti S, Alapati NS, Birajdar SS, et al. A scoping review of the effects of artificial intelligence on oral cancer treatment outcome and early diagnosis. Front Oral Maxillofac Med. 2025;7:4–4.

13. Prabhune AG, Srihari VR, Shree S, Katiyar M, Thampi V. Development and validation of an AI-based application for early detection and risk stratification of oral potentially malignant disorders. Oral 2025;15(6):1806–12.

14. Jung HA, Merkin R, Feng AL, Lee D, Lee K, Park S, et al. Artificial intelligence-powered real-time multimodal model for predicting recurrence and survival in head and neck cancer: a multicenter, multinational study. ESMO Open. 2026;11(2):106046.

15. Monani UJ, Yun TS, Sain M, Pattnaik PK. Improved classification of oral cancer through a personalized transfer learning CNN architecture. Oral 2025;15(6):1779–85.

16. Coole JB, Brenes D, Mitbander R, Vohra I, Hou H, Kortum A, et al. Multimodal optical imaging with real-time projection of cancer risk and biopsy guidance maps for early oral cancer diagnosis and treatment. J Biomed Opt. 2023;28(01):016002

17. Adeoye J, Su YX. Artificial intelligence for predicting post-excision recurrence and malignant progression in oral potentially malignant disorders: a retrospective cohort study. Int J Surg. 2026;112(1):1392–401.

18. Huang SY, Hsu RJ, Liu DW, Hsu WL. Using a machine learning algorithm and clinical data to predict the risk factors of disease recurrence after adjuvant treatment of advanced-stage oral cavity cancer. Tzu Chi Med J. 2025;37(1):91–8.

19. Tanriver G, Soluk Tekkesin M, Ergen O. Automated Detection and Classification of Oral Lesions Using Deep Learning to Detect Oral Potentially Malignant Disorders. Cancers. 2021;13(11):2766.

20. Flügge T, Gaudin R, Sabatakakis A, Tröltzsch D, Heiland M, Van Nistelrooij N, et al. Detection of oral squamous cell carcinoma in clinical photographs using a vision transformer. Sci Rep. 2023;13(1):2296.

21. Zhang J, Ding H, Zhu R, Liao W, Zhao J, Gao M, et al. gamUnet: designing global attention-based CNN architectures for enhanced oral cancer detection and segmentation. Front Med. 2025;12:1582439.

22. Lan T, Kuang S, Liang P, Ning C, Li Q, Wang L, et al. MRI-based deep learning and radiomics for prediction of occult cervical lymph node metastasis and prognosis in early-stage oral and oropharyngeal squamous cell carcinoma: a diagnostic study. Int J Surg. 2024;110(8):4648–59.

23. Aubreville M, Knipfer C, Oetter N, Jaremenko C, Rodner E, Denzler J, et al. Automatic Classification of Cancerous Tissue in Laserendomicroscopy Images of the Oral Cavity using Deep Learning. Sci Rep. 2017;7(1):11979.

24. Akram K, Aslam M, Waheed T, Ayesha N, Alamri FS, Mirdad AR, et al. Deep visual detection system for oral squamous cell carcinoma. Sci Rep. 2026;16(1):4219.

25. Shephard AJ, Bashir RMS, Mahmood H, Jahanifar M, Minhas F, Raza SEA, et al. A fully automated and explainable algorithm for predicting malignant transformation in oral epithelial dysplasia. npj Precis Onc. 2024;8(1):137.

26. Li J, Zhou Y, Zhang M, Adeoye J, Pu JJ, Zhou M, et al. Next-generation AI framework for comprehensive oral leukoplakia evaluation and management. npj Digit Med. 2025;8(1):513.

27. Shephard AJ, Mahmood H, Raza SEA, Damaceno Araujo AL, Santos-Silva AR, Lopes MA, et al. Development and Validation of an Artificial Intelligence-based Pipeline for Predicting Oral Epithelial Dysplasia Malignant Transformation [Internet]. Pathology; 2024 [cited 2026 Mar 10]. Available from: http://medrxiv.org/lookup/doi/10.1101/2024.11.13.24317264

28. Xu S, Liu Y, Hu W, Zhang C, Liu C, Zong Y, et al. An Early Diagnosis of Oral Cancer based on Three-Dimensional Convolutional Neural Networks. IEEE Access. 2019;7:158603–11.

29. Goswami B, Bhuyan MK, Alfarhood S, Safran M. Classification of Oral Cancer Into Pre-Cancerous Stages From White Light Images Using LightGBM Algorithm. IEEE Access. 2024;12:31626–39.

30. Devindi GAI, Dissanayake DMDR, Liyanage SN, Francis FBAH, Pavithya MBD, Piyarathne NS, et al. Multimodal Deep Convolutional Neural Network Pipeline for AI-Assisted Early Detection of Oral Cancer. IEEE Access. 2024;12:124375–90.

31. Haq IU, Ahmed M, Assam M, Ghadi YY, Algarni A. Unveiling the Future of Oral Squamous Cell Carcinoma Diagnosis: An Innovative Hybrid AI Approach for Accurate Histopathological Image Analysis. IEEE Access. 2023;11:118281–90.

32. Ramya S, Minu RI, Magesh KT. Xception Spiking Fractional Neural Network for Oral Squamous Cell Carcinoma Classification Based on Histopathological Images. IEEE Access. 2025;13:98574–91.

9. ANNEXES

9.1 Annex of a list of abbreviations

Table 3. List of abbreviations

Abbreviation	Meaning
AI	Artificial Intelligence
ML	Machine learning
DL	Deep Learning
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
OSCC	Oral squamous cell carcinoma
OC	Oral cancer
OPMD	Oral potentially malignant disorder
OED	Oral epithelial Dysplasia
NLP	Natural Language Processing
PICO	Population, Intervention, Comparison, Outcome
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
CLE	Confocal Laser Endomicroscopy
AUC	Area Under the Curve (of the ROC curve)
ROC	Receiver Operating Characteristic
F1-score	Harmonic mean of precision and recall
CT	Computed tomography