UMSHA Press

AJDR
Avicenna Journal of Dental Research

Avicenna J Dent Res 2026;18(2): 107-117. doi:10.34172/ajdr.4429

http://ajdr.umsha.ac.ir

Review Article

Using Artificial Intelligence in the Diagnosis, Classification, and Prognosis of Oral Squamous Cell Carcinoma: A Systematic Review and Meta-Analysis

Mohammad Mahdi Maleki[1*], Samaneh Vaziriamjad[2], Fatemeh Shahbazi[3]

[1]Student Research Committee, Hamadan University of Medical Sciences, Hamadan, Iran - Dental Implants Research Centre, Dental School, Hamadan University of Medical Sciences, Hamadan, Iran
[2]Department of Oral Medicine, School of Dentistry, Dental Research Center, Avicenna Institute of Clinical Sciences, Hamadan University of Medical Sciences, Hamadan, Iran
[3]Department of Epidemiology, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran

Article history:
Received: June 16, 2025
Revised: May 26, 2026
Accepted: May 31, 2026
ePublished: June 30, 2026

***Corresponding author:**
Mohammad Mahdi Maleki,
Email: M.m.maleki00@gmail.com

Abstract

Introduction: Oral squamous cell carcinoma (OSCC) is a prevalent malignancy within the head and neck region. Nonetheless, the overall prognosis remains poor, primarily due to the late-stage diagnosis in a substantial proportion of patients. Thus, this study aimed to critically evaluate the application of artificial intelligence (AI) in the diagnostic, classificatory, and prognostic processes of OSCC.

Methods: PubMed, Scopus, Web of Science and Embase databases were searched up to December 2024. Included studies reported relevant metrics (true positive, false positive, true negative, false negative, sensitivity, specificity, and area under the curve [AUC]). The review adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines.

Results: The search identified 1,863 articles, with 27 meeting the inclusion criteria. AI models achieved a pooled sensitivity of 0.90 (95% CI: 0.81–0.95), specificity of 0.89 (95% CI: 0.83–0.93), and an AUC of 0.733 (95% CI: 0.679–0.786). Diagnostic applications had the highest sensitivity, while classification tasks demonstrated superior specificity. Ultimately, prognostic models displayed lower sensitivity and specificity, indicating variability in performance across different AI applications.

Conclusion: Rapid development of AI algorithms has facilitated their application in OSCC diagnosis, classification, and prognosis. Additionally, AI models have shown high sensitivity and specificity in diagnostic applications, promising earlier detection and improved classification accuracy, which can lead to superior therapeutic outcomes. Moreover, their non-invasive nature compared to histopathological biopsy enhances patient comfort while reducing costs. Overall, AI holds significant potential to transform OSCC diagnosis and prognosis, finally affecting community oral health.

Keywords: Artificial intelligence, Oral squamous cell carcinoma, Diagnosis, Prognosis, Classification

Please cite this article as follows: Maleki MM, Vaziriamjad S, Shahbazi F. Using artificial intelligence in the diagnosis, classification, and prognosis of oral squamous cell carcinoma: a systematic review and meta-analysis. Avicenna J Dent Res 2026;18(2): 107-117. doi:10.34172/ajdr.4429

Introduction

Oral cancer ranks as the sixteenth most prevalent cancer globally, with significant morbidity and mortality rates (1). Oral squamous cell carcinoma (OSCC) arises from the oral mucosa and is considered a common malignancy in the head and neck region (2). It should be noted that OSCC predominantly affects men more than women and is frequently observed in middle-aged and older adults (3). Moreover, it can result in severe facial disfigurement and impairments in swallowing, speech, and taste, thereby severely impacting patients' quality of life (4, 5).

Clinically, OSCC typically presents as either a red and white lesion or a completely red lesion with an irregular surface (6, 7). In addition, early-stage lesions are often painless (8), but as the disease progresses, they can become painful and may exhibit features such as ulceration, nodularity, and attachment to surrounding tissues (9). Ulceration is a common sign of OSCC, which is characterized by an irregular surface and borders, with the lesion's edges feeling firm upon palpation (9, 10). Further, the lateral and posterior surfaces of the tongue are the most frequent sites for OSCC, accounting for approximately 50% of cases (11), though other areas (e.g., the floor of the mouth, soft palate, gingiva, buccal mucosa, and hard palate) are

normally affected as well (12). Similarly, OSCC primarily spreads via lymphatic drainage to the ipsilateral cervical lymph nodes, but it can also metastasize to contralateral or bilateral nodes. Distant metastases are typically observed in the lungs, bones, and liver (13).

Likewise, OSCC is a highly aggressive form of oral cancer, and delays in diagnosis can significantly worsen prognosis (14). The diagnosis of OSCC involves correlating clinical features with histopathological findings, with more than 90% of cases diagnosed morphologically as OSCC (15). However, challenges in diagnosis may arise due to discrepancies between clinical presentations and histopathological interpretations (16).

The early diagnosis of OSCC is essential for improving treatment outcomes and increasing survival rates while reducing both morbidity and mortality (17). While the histopathological examination of tissue samples remains the gold standard for diagnosing and grading oral cancer, it is prone to errors (18). Furthermore, the variability in interpretations can influence clinical decisions and treatment strategies (19). As a result, there is an increasing need for alternative diagnostic approaches that offer greater precision, faster results, and more consistency. Recently, significant strides have been made in harnessing the power of artificial intelligence (AI) to enhance diagnostic accuracy. Notably, Landini and Osman pioneered the development of an automated approach using morphological reconstruction in order to identify epithelial characteristics for diagnosing oral cancer (20). Additionally, digital pathology, which integrates advanced computer technologies, is emerging as a robust tool for high-throughput, quantitative analysis in the field of cancer diagnosis (21-23).

Machine-learning (ML) techniques are employed to detect patterns within existing datasets, though they rely on human expertise to manually define and extract key features. On the other hand, deep learning (DL), a specialized subset of ML, utilizes artificial neural networks to replicate brain-like learning processes, thereby allowing the system to autonomously extract features from raw image data. Both ML and DL have significantly enhanced the capabilities of computer-aided diagnostic systems, especially with the increase in training data (24). Researchers have combined various methodologies, including image processing, pattern recognition, ML, and DL, to develop advanced computer-aided diagnostic systems for OSCC diagnosis. The automation of oral cancer quantification has proven effective in minimizing discrepancies among pathologists, ensuring more accurate and consistent diagnoses (25).

Numerous studies have investigated the application of AI in the early diagnosis, prognosis, and classification of OSCC. This systematic review and meta-analysis will also evaluate the performance of AI in diagnosing, classifying, and predicting OSCC.

Materials and Methods

Protocol and Registration

This systematic review and meta-analysis was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.

Search Strategy

Four electronic databases, including PubMed, Scopus, Embase, and Web of Science, were comprehensively searched using the following search strategy: ((((((Artificial Intelligence[MeSH Terms]) OR (Computational Intelligence[Title/Abstract])) OR (Machine Intelligence[Title/Abstract])) OR (Machine Intelligence[MeSH Terms])) OR (Deep Learning[MeSH Terms])) AND (((((((Head and Neck Neoplasms[MeSH Terms]) OR (Head[Title/Abstract] AND Neck Cancer[Title/Abstract])) OR (Precancerous Conditions[MeSH Terms])) OR (Neck Cancers[Title/Abstract])) OR (Neck Cancer[Title/Abstract])) OR (Head Cancers[Title/Abstract])) OR (Head Cancer[Title/Abstract]))) AND ((((Diagnosis[MeSH Terms]) OR (Early Diagnosis[MeSH Terms])) OR (Early Detection of Disease[Title/Abstract])) OR (Disease Early Detection[Title/Abstract])) up to December 2024. All studies identified through the search were imported into EndNote software (version 20) for further processing, where duplicates were detected and removed. Then, the reference lists of eligible articles were manually reviewed to ensure the inclusion of all relevant studies. It should be noted that the search was not limited by language or publication date.

Eligibility Criteria

The inclusion criteria were predefined, focusing on studies that demonstrated the application of AI in the diagnosis, classification, or prognosis of OSCC. Eligible studies were required to report metrics such as true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), or provide data enabling the calculation of sensitivity (Se), specificity (Sp), or area under the curve (AUC). Studies with AI models capable of distinguishing OSCC from other tissue types, whether benign, precancerous, or cancerous, were included as well. However, studies failing to meet these criteria were excluded from the investigation. Likewise, studies involving animal models, as well as systematic reviews, meta-analyses, clinical trials, and randomized controlled trials, were excluded from this review. When multiple articles used the same dataset, only the study with the larger sample size was retained for analysis.

Study Selection and Screening

The remaining studies were imported into EndNote Version 20 for further screening. Two reviewers independently assessed titles and abstracts, and any disagreements were resolved by consensus. Furthermore, full-text versions of studies deemed eligible were retrieved and independently screened twice. Discrepancies were also resolved through group discussion.

Data Extraction

Data from each eligible study were extracted, including the first author's name, publication date, study design,

participant demographics (age and gender), country of origin, AI index test, reference test, and diagnostic outcomes (i.e., TP, FP, TN, FN). Additionally, sensitivity, specificity, AUC, training and validation datasets, and OSCC prevalence were recorded. In addition, diagnostic accuracy metrics for multiple AI models within a single study were recorded separately. The highest sensitivity and specificity were selected if a dataset was validated multiple times for a model. This situation typically occurs when a dataset is used in repeated validation procedures, such as cross-validation, where the data are split into different subsets for training and testing multiple times. Sensitivity refers to the model's ability to correctly identify positive cases, while specificity represents its ability to correctly identify negative cases. Two researchers independently completed this task, with any disagreements resolved by a third reviewer. Finally, all extracted data were compiled in a Microsoft Excel file (version 2019), which was accessible to all authors.

Quality of Studies

The methodological quality of the selected studies was evaluated using the Quality Assessment Tool for Diagnostic Accuracy Studies (QUADAS-2) (26). This assessment was independently conducted by two researchers, with group discussions held when necessary. The focus was on the safeguards against bias in each study and, rather than categorizing bias risk, assigned a numerical value to the safeguards: a value of 1 for safeguards implemented ("yes/low bias") and 0 for those not implemented ("no/high bias/unclear"). This process allowed us to apply a bias-adjusted meta-analysis model through a quantitative approach (27, 28). To reduce subjectivity in the QUADAS-2 assessment, the judgment question in each domain was omitted, focusing on the presence or absence of safeguard measures. While this approach increases reproducibility and simplifies the bias assessment process, it may overlook nuanced aspects of potential bias that could be relevant for interpretation. Ultimately, this method resulted in fourteen comprehensive bias safeguards (Figure 1).

Data Analysis

For data synthesis, diagnostic information was extracted directly from each study, including the number of TP, FP, TN, and FN. Sensitivity and specificity with 95% confidence intervals were used when studies did not report 2 × 2 tables explicitly. In cases where multiple validations were reported for the same dataset, the highest reported sensitivity and specificity were selected for meta-analysis. Although this approach allows the inclusion of the most favorable estimates, it may introduce upward bias in pooled results. Accordingly, future studies could consider hierarchical or multilevel modeling to account for repeated validations within the same dataset.

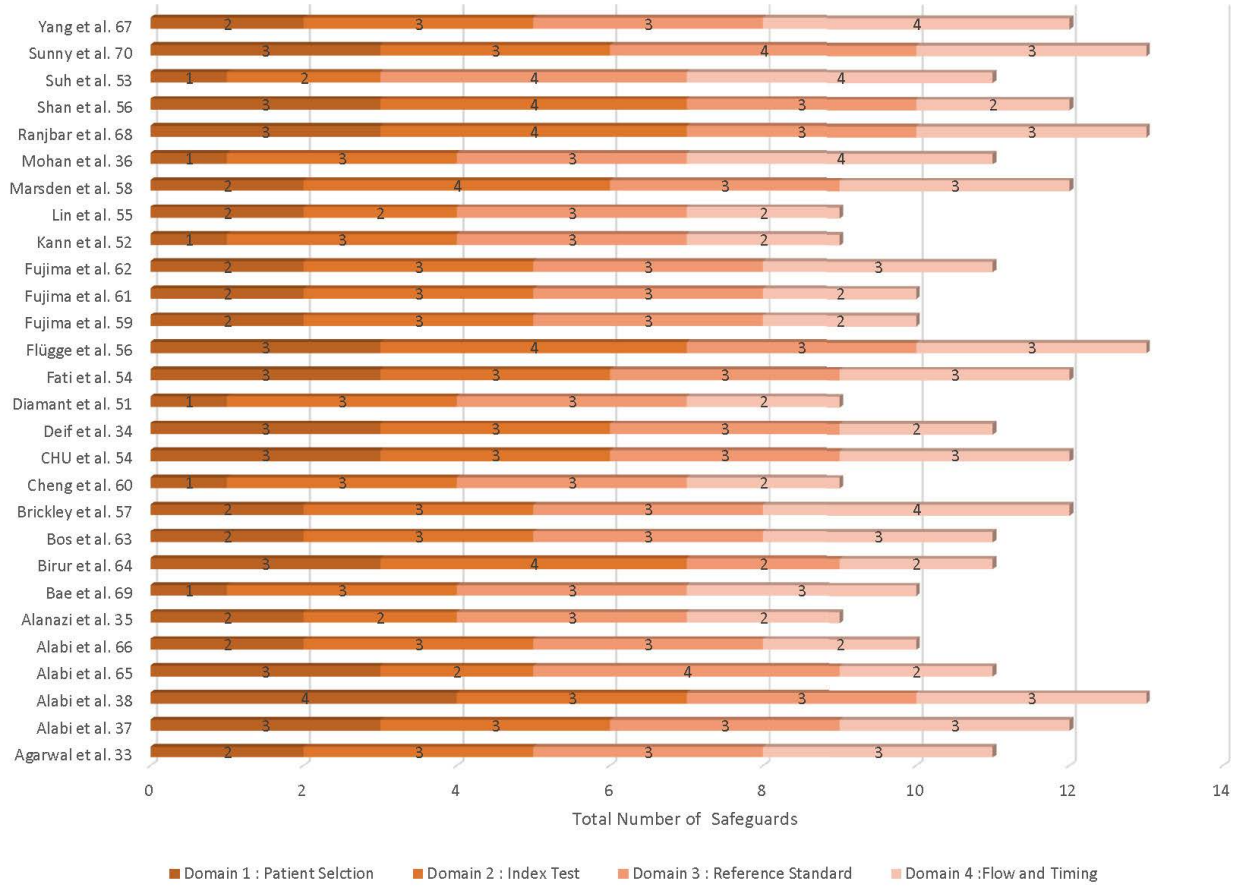

Figure 1. Quality Assessment Using QUADAS-2 for the Included Studies
Note. QUADAS-2: The Quality Assessment Tool for Diagnostic Accuracy Studies

Using MedCalc software (version 20), the standard errors for each diagnostic effect size were calculated based on the extracted data. MedCalc employs methods suitable for 2 × 2 diagnostic tables to estimate pooled sensitivity, specificity, and AUC) while accounting for between-study variability. Moreover, a random-effects model was applied to combine the results across studies, allowing for heterogeneity in diagnostic performance. Additionally, between-study heterogeneity was assessed using the I^2 statistic.
Likewise, publication bias was evaluated using both Egger's and Begg's tests. All graphical representations, including forest plots and summary receiver operating characteristic curves, were generated using MedCalc, version 20 (29, 30).

Subgroup Analysis

To better assess AI applications, studies were divided into two approaches. First, studies were categorized into diagnosis, classification, and prognosis, with each group separately analyzed to evaluate AI performance by function. Diagnosis refers to AI models that identify the presence or absence of a specific disease. In addition, classification involves categorizing diseases or tissue types based on their distinct characteristics. Further, prognosis focuses on predicting patient outcomes or treatment responses. Each of these categories was separately analyzed to comprehensively assess AI performance. In the second classification, studies were assigned to computed tomography (CT) scans, clinical and pathologic records, photography, magnetic resonance (MR), positron emission tomography, and histopathology groups, with each category separately analyzed due to methodological homogeneity.

Results

Study Selection

The PRISMA flowchart in Figure 2 illustrates the study selection process. Our comprehensive database search initially identified 1,863 articles. After de-duplication, 610 were removed via EndNote 20, leaving 1,253 for title and abstract screening. After title and abstract review, 1,142 publications were excluded, and only 111 remained for full-text review. Ultimately, due to the meta-analysis requirement, 27 articles with data on sensitivity, specificity, or AUC were included in the analysis.

Study and Index Test Characteristics

Table 1 provides the datasets extracted from the included articles. They were published between 1996 and 2023, mostly from the United States (n = 6) and India (n = 5). The most common test was histopathology (n = 7), followed by clinical and pathologic records (n = 6), CT (n = 3), positron emission tomography (PET, n = 4), photography (n = 4), and MR (n = 3).

Main Analysis Results

Twenty-eight datasets, including 39,204 samples, underwent analysis. Overall, AI had a sensitivity of 0.90 (0.81–0.95), a specificity of 0.89 (0.83–0.93), and an AUC of 0.733 (95% CI: 0.679–0.786). In the analysis of AI functions (diagnosis, classification, and prognosis), 13 studies assessed AI for SCC diagnosis, showing a sensitivity of 0.92 (0.84–0.97), a specificity of 0.89 (0.80–0.94), and an AUC of 0.785 (95% CI: 0.712–0.857). For disease classification, five studies were analyzed, with a sensitivity of 0.89 (0.79–0.94) and a specificity of 0.90 (0.79–0.95); AUC was not reported due to the absence of studies reporting it. Nine studies evaluated prognosis, displaying a sensitivity of 0.48 (0.06–0.91), a specificity of 0.80 (0.60–0.92), and an AUC of 0.674 (95% CI: 0.652–0.696) (Figures 3A-D and 4).

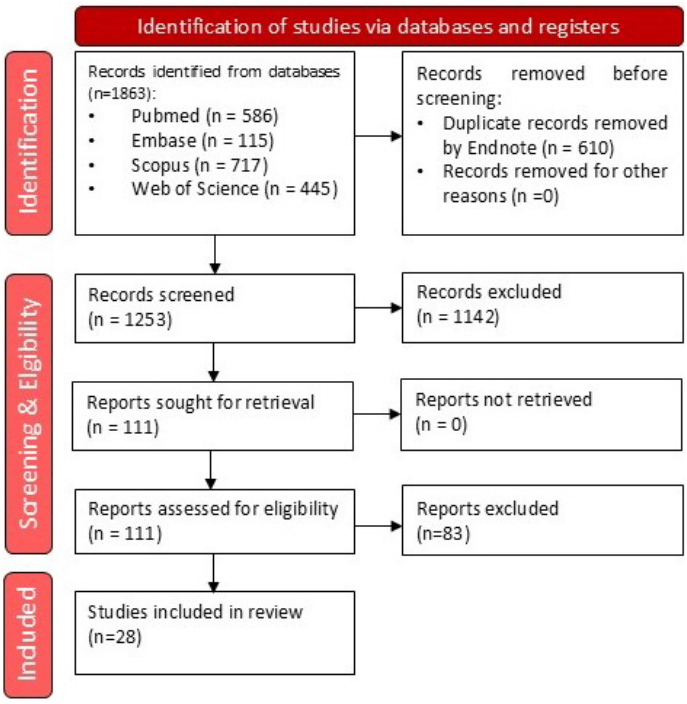

Figure 2. PRISMA Flowchart
Note. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Table 1. Characteristics and Extracted Data From Included Studies

First Author (Year)	Country	Type of Process	Index Test	Samples (n)	Training (n)	Validation (n)	Sensitivity	Specificity	AUC	Type of ML Model
Adwan A. Alanazi (2022) (31)	Saudi Arabia	Classifying	Photography	NR	90%	10%	0.94	0.94	NR	Deep convolutional neural network (CNN)
André Diamant (2019) (32)	Canada	Prognosis	CT	300	194	106	NR	NR	0.666 (0.636–0.696)	DL (fully connected neural network)
Benjamin H Kann (2023) (33)	United States of America	Diagnosis	CT	360	NR	NR	NR	NR	0.720 (0.670–0.770)	DL (CNN)
Chong Hyun Suh (2020) (34)	Republic of Korea	Diagnosis	MR	60	40	20	NR	NR	0.744 (0.611–0.877)	ML (radiomics + SVM)
Chui Shan CHU (2020) (35)	Hong Kong	Prognosis	Histopathology	408 (467)	NR	NR	0.30	0.87	NR	ML (gradient boosting/decision tree)
Huiping Lin (2021) (36)	China	Diagnosis	Photography	760	NR	NR	NR	NR	0.946 (0.922–0.970)	DL (CNN-based mobile image classifier)
Jie Shan (2020) 7(37)	China	Diagnosis	Clinical and pathologic records	101	70%	30%	0.82	0.78	0.773 (0.685–0.861)	ML (logistic regression/SVM)
M. R. Brickley (1996) (38)	United Kingdom	Diagnosis	Clinical and pathologic records	348	223	81	0.81	0.80	NR	Classical statistical model (rule-based classifier)
Mark Marsden (2020) (39)	United States of America	Diagnosis	Histopathology	53	NR	NR	NR	NR	0.738 (0.712–0.764)	ML (RF classifier)
Mohanad A. Deif (2022) (40)	India	Diagnosis	Histopathology	1224	80%	20%	0.98	0.83	NR	Deep neural network + binary particle swarm optimization
Noriyuki Fujima (2021) (41)	United States of America	Prognosis	PET	99	66	33	NR	NR	0.656 (0.583–0.729)	DL (CNN)
Nai-Ming Cheng (2021) (42)	Taiwan	Prognosis	PET	NR	NR	NR	NR	NR	0.622 (0.375–0.869)	DL (fully automated CNN)
Noriyuki Fujima (2020) (43)	United States of America	Prognosis	PET	120	90	30	NR	NR	0.674 (0.619–0.729)	DL (CNN model)
Noriyuki Fujima (2021) (44)	United States of America	Prognosis	PET	154	102	52	NR	NR	0.706 (0.657–0.755)	DL (CNN model)
Paula Bos (2021) (45)	Netherlands	Prognosis	MR	177	70%	30%	NR	NR	0.642 (0.283–1.000)	ML (logistic regression/RF)
Praveen Birur N (2022) (46)	India	Diagnosis	Photography	13490	10658	2832	0.73	0.78	NR	DL (MobileNet CNN for point-of-care device)
Ramya Mohan (2023) (47)	India	Classifying	Histopathology	1214	NR	NR	0.96	0.97	NR	DL (fused CNN framework)
Rasheed Omobolaji Alabi (2022) (48)	Finland	Classifying	Clinical and pathologic records	3284	3164	120	0.89	0.91	NR	ML (RF, logistic regression, and SVM)
Rasheed Omobolaji Alabi (2019) (49)	Finland	Classifying	Clinical and pathologic records	311	50%	50%	0.82	0.71	NR	ML (logistic regression, and SVM)
Rasheed Omobolaji Alabi (2019) (50)	Finland	Prognosis	Clinical and pathologic records	311	50%	50%	0.17	0.55	NR	ML (RF, and SVM)
Rasheed Omobolaji Alabi (2021) (51)	Finland	Prognosis	Clinical and pathologic records	7699	7649	50	1.00	0.87	NR	ML (nomogram and ML comparison Models)
S.Y. Yang (2022) (52)	China	Diagnosis	Histopathology	2124	1925	199	0.95	0.94	NR	DL (CNN)
Sara Ranjbar (2017) (53)	United States of America	Classifying	CT	107	NR	NR	0.75	0.80	NR	Radiomics + ML (SVM)
Sohi Bae (2020) (54)	Republic of Korea	Diagnosis	MR	174	NR	NR	NR	NR	0.75 (0.613–0.887)	Radiomics-based ML (RF/SVM)
Suliman Mohamed Fati (2022) (55)	Saudi Arabia	Diagnosis	Histopathology	5192	80%	20%	0.97	0.98	NR	DL (hybrid CNN + transfer learning)

Table 1. Continued.

First Author (Year)	Country	Type of Process	Index Test	Samples (n)	Training (n)	Validation (n)	Sensitivity	Specificity	AUC	Type of ML Model
Sumsum Sunny (2019) (56)	India	Diagnosis	Histopathology	100	NR	NR	0.73	0.81	NR	ML (deep CNN for Tele-cytology platform)
Tabea Flügge (2023) (57)	Germany	Diagnosis	Photography	1406	1265	141	0.99	0.99	NR	Vision transformer (DL model)

Note. SVM: Support vector machine; AUC: Area under the curve; ML: Machine learning; DL: Deep learning; RF: Random forest; CT: Computed tomography; MR: Magnetic resonance; NR: Not reported

In terms of tests, photography and histopathology had the best sensitivity and specificity among all test categories. Photography had a sensitivity of 0.93 (0.72–0.99) and a specificity of 0.94 (0.76–0.99), while histopathology had a sensitivity of 0.91 (0.72–0.98) and a specificity of 0.93 (0.76–0.97). Other laboratory methods were ranked lower. Moreover, studies using clinical and pathologic records had a sensitivity of 0.78 (0.57–0.91) and a specificity of 0.80 (0.68–0.88). AUCs in studies using CT, MR, and PET were reported as 0.694 (95% CI: 0.650–0.737), 0.689 (95% CI: 0.622–0.756), and 0.661 (95% CI: 0.630–0.692), respectively (Figures 3E-J and 5).

Quality Assessment and Publication Bias

The methodological quality of the diagnostic accuracy studies selected for inclusion was assessed using the QUADAS-2 tool. This assessment was independently conducted by two reviewers, with any discrepancies addressed through group discussions when necessary. The tool specifically examines potential biases related to disease progression, partial verification, differential verification, incorporation, test review, diagnostic review, and clinical review. Studies without any evidence of these biases were considered to have a low risk of bias. The Begg's ($P=0.2726$) and Egger's ($P=0.2785$) tests indicated no significant publication bias in the use of AI for diagnosing, classifying, and predicting the prognosis of OSCC lesions (Figure 1).

Discussion

Principal Findings

With the rapid advancement of computer algorithms, AI is increasingly used for diagnosing, classifying, and predicting the prognosis of OSCC patients using various methods. This systematic review and meta-analysis study assessed the diagnostic accuracy of various AI models and demonstrated their high accuracy in diagnosing, classifying, and predicting OSCC prognosis in patients. In the study by Flügge et al, the highest diagnostic sensitivity and specificity were observed using a vision transformer for OSCC detection in clinical photographs (57). Similarly, the study performed by Fati et al (55) for early OSCC diagnosis using histology images confirmed high sensitivity and specificity. However, a study conducted by Deif et al (40) using CT demonstrated lower specificity despite high sensitivity. For OSCC classification, Alanazi et al and Mohan et al using biomedical images and histopathology images reported the highest sensitivity and specificity, with subsequent studies following closely (31, 47). There are fewer studies regarding the use of AI in OSCC prognosis assessment. Likewise, Alabi et al, investigating SCC patients, reported the highest sensitivity and specificity (50, 51). Given the limited number of studies on OSCC prognosis with AI and the importance of predicting lesion and patient outcomes for early treatment and alternative therapeutic strategies, further studies are needed to explore AI in this area.

Other Applications of Artificial Intelligences in Head and Neck Cancers

Today, the use of AI in medicine and the assessment of patients and related lesions has increased with its increasing advancement in various fields and the high significance of medicine and health. Due to the high risk and recurrence of cancer, AI has received more attention than before. In addition to oral cancers, head and neck cancers are of great importance due to the vital proximity of the area and high prevalence (58-62). In head and neck oncology, ML methods have been widely investigated in all imaging techniques, including ultrasound, CT, MR, and nuclear medicine (63, 64). Using these tools, by standardizing the opinions of various researchers, reduces discrepancies among pathologists (25). Moreover, several studies have been conducted on AI use in surgery, including decision-making, outcome prediction, and surgical approach development, confirming the value of AI in head and neck cancer surgery (65-68). However, knowledge about AI and its potential applications remains scarce among otolaryngologists and surgeons, who are critical in primary care for H&N cancers, requiring further studies and exploration (69).

Issues in Current Practice

Although biopsy, microscopic examination, and histopathological analysis of oral tissues are the gold standards for cancer evaluation, other auxiliary techniques can be employed to assess lesions. However, sampling errors can occur due to its invasive nature, potential sampling or transfer errors, and the lesion's non-uniformity. Additionally, the shortage of pathologists and the reliance solely on human intelligence, due to inherent errors, can make lesion evaluation more challenging. Lack of access to necessary facilities and specialists and human errors are recurring issues in other lesion assessment methods.

Considering that early diagnosis and disease classification can lead to quicker, less invasive treatments and better patient prognoses, AI can fulfill this need. Additionally, AI can provide greater accuracy through various assessment methods. Furthermore, due to its reduced reliance on

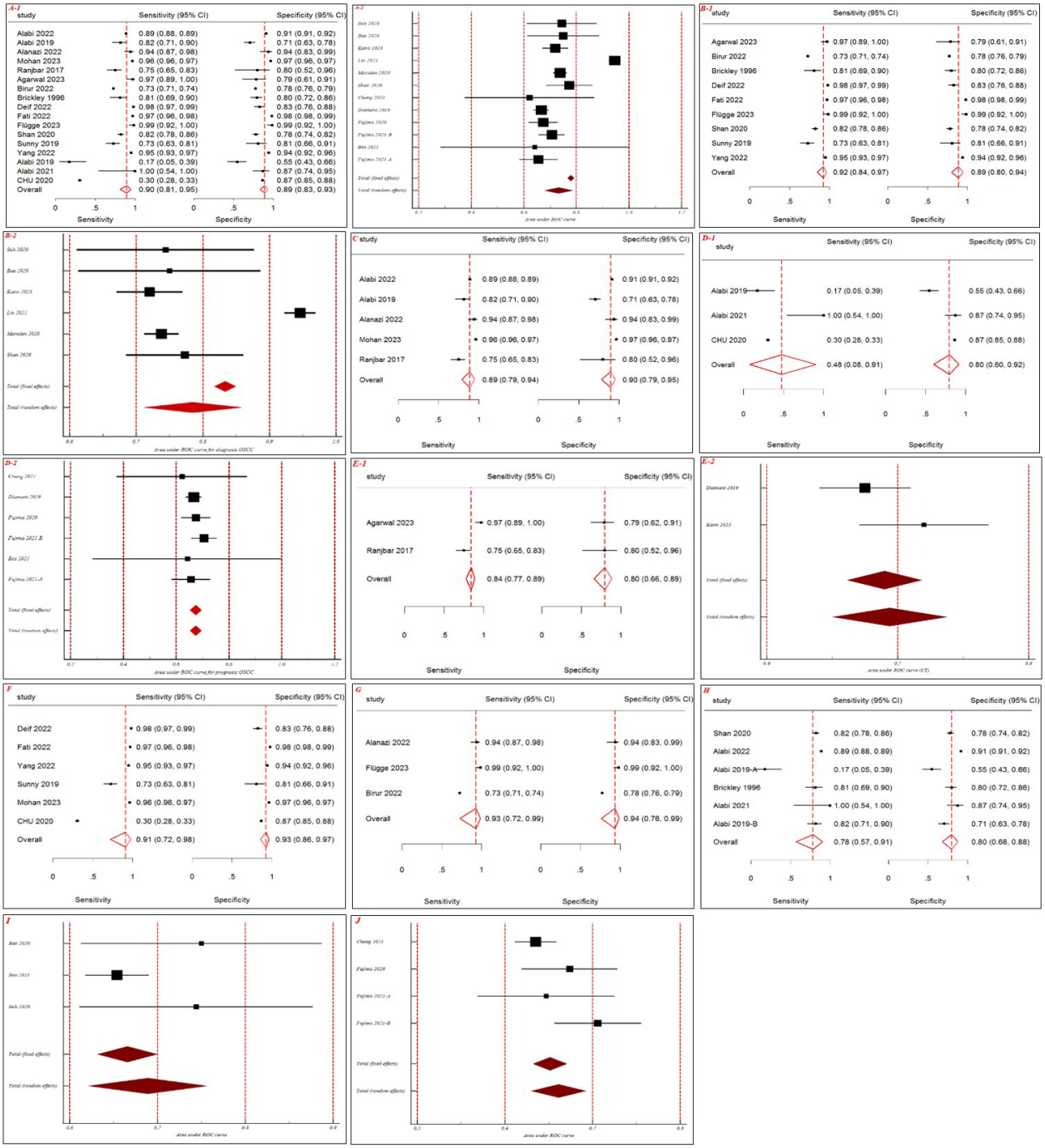

Figure 3. The results of Analysis of Studies in Different Categories: Total **(A)**, Type of Process, Diagnosis **(B)**, Classifying **(C)**, Prognosis **(D)**, Index Test, CT **(E)**, Histopathology **(F)**, Photography **(G)**, Clinical and Pathologic Records **(H)**, MR **(I)**, and PET **(J)**
Note. CT: Computed tomography; MR: Magnetic resonance; PET: Positron emission tomography. All parts of this figure are created by the authors

specialist expertise, AI can offer faster and broader access to results. In addition, it can be an auxiliary tool to help specialists diagnose and evaluate more accurately, thereby reducing the likelihood of human error while increasing their precision.

Limitations

This systematic review and meta-analysis had certain limitations. Initially, more studies were identified and reviewed, but due to the lack of criteria required for analysis, they could not be included in our meta-analysis. It should be noted that the number of studies assessing AI for the prognosis of OSCC was relatively small. This limited sample size likely contributed to the lower pooled sensitivity and specificity observed in this subgroup. Consequently, the robustness of conclusions regarding prognostic performance is somewhat limited, and further high-quality studies are needed to strengthen evidence in this area. Likewise, many of the included studies provided limited details on AI model validation. Variation in data splitting and validation methods (e.g., cross-validation or external validation) may affect results and introduce potential over-fitting. Thus, future studies should report these processes more clearly in order to improve reliability

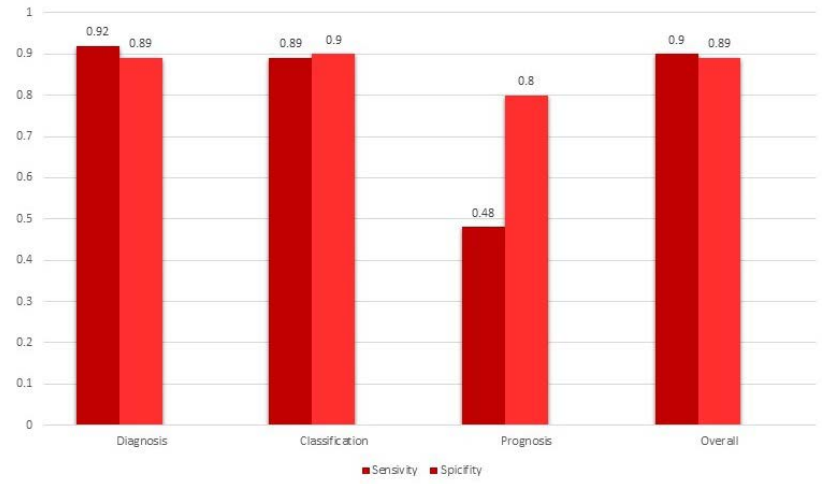

Figure 4. Graphical Comparison of Analysis's Results of the Included Studies in Various Categories: Total, Type of process, Diagnosis, Classifying, and Prognosis

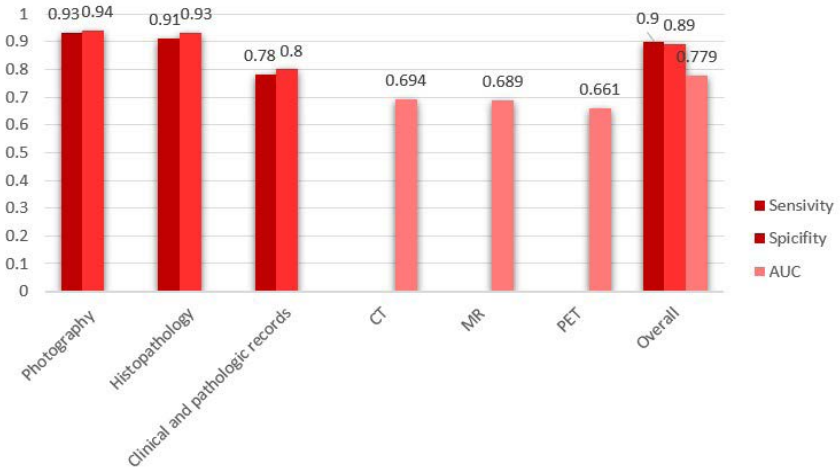

Figure 5. Graphical Comparison of Analysis's Results of Included Studies in Varied Categories

and generalizability. Additionally, although the included studies reported diagnostic accuracy metrics (e.g., sensitivity, specificity, and AUC), limited information was available on clinical utility measures, (e.g., positive predictive value, negative predictive value, or decision curve analysis). Consequently, the real-world impact of AI-assisted diagnosis on patient outcomes remains unclear. Hence, future studies should incorporate these metrics to better assess the clinical relevance of AI in OSCC management.

Likewise, our approach to QUADAS-2, which omitted the judgment question, may have oversimplified the assessment of bias. Some subtle but potentially important aspects of study quality might not have been captured, and this should be considered when interpreting the pooled results. While our systematic review and meta-analysis synthesized available AI studies for OSCC, many of the included studies did not completely adhere to AI-specific reporting standards (e.g., CONSORT-AI or TRIPOD-AI). Incorporating these standards in future research would improve the transparency, reproducibility, and comparability of AI diagnostic and prognostic studies. Ultimately, some included studies, particularly those evaluating AI for classification tasks, did not report AUC values, limiting a comprehensive comparison of diagnostic performance across different AI applications and highlighting the need for complete reporting of all relevant accuracy metrics in future research.

Policy Implications and Future Research

Datasets other than those used for training should be utilized for a better assessment of AI performance. To ensure more capable AI, diverse datasets should be employed for training, making the system more reliable. To achieve such machines, a database should first be established, allowing various machines to be trained on broader datasets as needed. In addition, studies in AI usage should report more data, including TP, FP, TN, FN, sensitivity, and specificity, for comparative studies.

Further, AI-based screening of OSCC lesions can be routinely conducted in large populations, preventing disease progression with early diagnosis and leading to better prognosis with more superficial treatment. AI use can reduce specialists' human error and analyze patient

data to arrive at more accurate conclusions, especially in cases of disagreement. Furthermore, AI can assist in determining OSCC prognosis since human intelligence cannot explore all data and reach a specific prognosis, but AI can be a guide for necessary treatment. However, studies on OSCC prognosis with AI are limited, and given its importance in treatment selection, it is recommended that future studies focus more on this area.

Another recommendation is to conduct more studies using multiple methods simultaneously, as this approach can yield more accurate results based on more comprehensive data and provide more robust and broader findings.

Conclusion

The findings revealed that early diagnosis of OSCC plays a crucial role in treatment and improvement of disease prognosis. Moreover, the staging and grading of the disease significantly could help in treatment planning and prognosis prediction. Using AI as an auxiliary tool for probable OSCC diagnosis was found to significantly assist in speeding up diagnosis and determining disease staging and grading. In addition, the data demonstrated this tool can help with accurate and faster diagnosis and can be more accessible, cheaper, and more manageable, allowing for broader community use, even for screening purposes. Based on the findings, AI can use various methods to achieve more accurate and acceptable results. Additionally, compared to the current gold standard of biopsy for histopathology tests, it is non-invasive and more acceptable to patients, with lower costs. However, ethical and legal concerns in this area require greater attention. Finally, AI can also be utilized as an auxiliary tool in the early detection of secondary OSCCs and the recurrence of lesions.

Authors' Contribution

Conceptualization: Mohammad Mahdi Maleki, Samaneh Vaziriamjad
Data Curation: Mohammad Mahdi Maleki
Formal Analysis: Fatemeh Shahbazi
Funding Acquisition: Samaneh Vaziriamjad
Investigation: Mohammad Mahdi Maleki
Methodology: Samaneh Vaziriamjad, Mohammad Mahdi Maleki, Fatemeh Shahbazi
Project Administration: Mohammad Mahdi Maleki
Resources: Mohammad Mahdi Maleki
Software: Fatemeh Shahbazi
Supervision: Samaneh Vaziriamjad
Validation: Fatemeh Shahbazi
Visualization: Mohammad Mahdi Maleki
Writing–Original Draft: Mohammad Mahdi Maleki
Writing–Review & Editing: Samaneh Vaziriamjad, Mohammad Mahdi Maleki

Consent for Publication

All the authors of this paper affirm that it has been solely submitted to this journal and that it is not concurrently under consideration for publication in another journal. Additionally, all the named authors have been involved in the work leading to the publication of the paper and have read and confirmed the paper before its submission for publication.

Data Availability Statement

The datasets used and/or analyzed during the current study are available upon reasonable request from the corresponding author.

Ethical Approval

This study is based solely on secondary data extracted from previously published studies in public databases. No direct involvement of human participants or identifiable patient data occurred accordingly. Therefore, ethical approval was not applicable or necessary.

Funding

This study was approved and financially supported by the Research Center of Hamadan University of Medical Sciences (grant No. 140303222296).

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71(3):209-49. doi:10.3322/caac.21660
2. Mody MD, Rocco JW, Yom SS, Haddad RI, Saba NF. Head and neck cancer. Lancet 2021;398(10318):2289-99. doi:10.1016/s0140-6736(21)01550-6
3. Safi AF, Kauke M, Grandoch A, Nickenig HJ, Drebber U, Zöller J, et al. Clinicopathological parameters affecting nodal yields in patients with oral squamous cell carcinoma receiving selective neck dissection. J Craniomaxillofac Surg 2017;45(12):2092-6. doi:10.1016/j.jcms.2017.08.020
4. Linsen SS, Gellrich NC, Krüskemper G. Age- and localization-dependent functional and psychosocial impairments and health related quality of life six months after OSCC therapy. Oral Oncol 2018;81:61-8. doi:10.1016/j.oraloncology.2018.04.011
5. Meier JK, Schuderer JG, Zeman F, Klingelhöffer C, Hullmann M, Spanier G, et al. Health-related quality of life: a retrospective study on local vs. microvascular reconstruction in patients with oral cancer. BMC Oral Health 2019;19(1):62. doi:10.1186/s12903-019-0760-2
6. Maymone MB, Greer RO, Kesecker J, Sahitya PC, Burdine LK, Cheng AD, et al. Premalignant and malignant oral mucosal lesions: Clinical and pathological findings. J Am Acad Dermatol 2019;81(1):59-71. doi:10.1016/j.jaad.2018.09.060
7. Odell E, Kujan O, Warnakulasuriya S, Sloan P. Oral epithelial dysplasia: recognition, grading and clinical significance. Oral Dis 2021;27(8):1947-76. doi:10.1111/odi.13993
8. Brandizzi D, Gandolfo M, Velazco ML, Cabrini RL, Lanfranchi HE. Clinical features and evolution of oral cancer: a study of 274 cases in Buenos Aires, Argentina. Med Oral Patol Oral Cir Bucal 2008;13(9): E544-8.
9. Bagan J, Sarrion G, Jimenez Y. Oral cancer: clinical features. Oral Oncol 2010;46(6):414-7. doi:10.1016/j.oraloncology.2010.03.009
10. Scully C, Bagan J. Oral squamous cell carcinoma: overview of current understanding of aetiopathogenesis and clinical implications. Oral Dis 2009;15(6):388-99. doi:10.1111/j.1601-0825.2009.01563.x
11. Harada H, Kikuchi M, Asato R, Hamaguchi K, Tamaki H, Mizuta M, et al. Characteristics of oral squamous cell carcinoma focusing on cases unaffected by smoking and drinking: a multicenter retrospective study. Head Neck 2023;45(7):1812-22. doi:10.1002/hed.27398
12. Omura K. Current status of oral cancer treatment strategies: surgical treatments for oral squamous cell carcinoma. Int J Clin Oncol 2014;19(3):423-30. doi:10.1007/s10147-014-0689-z
13. Jerjes W, Upile T, Petrie A, Riskalla A, Hamdoon Z, Vourvachis M, et al. Clinicopathological parameters, recurrence, locoregional and distant metastasis in 115 T1-T2 oral squamous cell carcinoma patients. Head Neck Oncol 2010;2:9. doi:10.1186/1758-3284-2-9
14. McCullough MJ, Prasad G, Farah CS. Oral mucosal malignancy and potentially malignant lesions: an update on

the epidemiology, risk factors, diagnosis and management. Aust Dent J 2010;55 Suppl 1:61-5. doi:10.1111/j.1834-7819.2010.01200.x
15. Tandon P, Dadhich A, Saluja H, Bawane S, Sachdeva S. The prevalence of squamous cell carcinoma in different sites of oral cavity at our Rural Health Care Centre in Loni, Maharashtra - a retrospective 10-year study. Contemp Oncol (Pozn) 2017;21(2):178-83. doi:10.5114/wo.2017.68628
16. Elmakaty I, Elmarasi M, Amarah A, Abdo R, Malki MI. Accuracy of artificial intelligence-assisted detection of oral squamous cell carcinoma: a systematic review and meta-analysis. Crit Rev Oncol Hematol 2022;178:103777. doi:10.1016/j.critrevonc.2022.103777
17. Chakraborty D, Natarajan C, Mukherjee A. Advances in oral cancer detection. Adv Clin Chem 2019;91:181-200. doi:10.1016/bs.acc.2019.03.006
18. Santana MF, Ferreira LC. Diagnostic errors in surgical pathology. J Bras Patol Med Lab 2017;53(2):124-9. doi:10.5935/1676-2444.20170021
19. Kong J, Sertel O, Shimada H, Boyer KL, Saltz JH, Gurcan MN. Computer-aided evaluation of neuroblastoma on whole-slide histology images: classifying grade of neuroblastic differentiation. Pattern Recognit 2009;42(6):1080-92. doi:10.1016/j.patcog.2008.10.035
20. Landini G, Othman IE. Architectural analysis of oral cancer, dysplastic, and normal epithelia. Cytometry A 2004;61(1):45-55. doi:10.1002/cyto.a.20082
21. Warnke-Sommer JD, Ali HH. Evaluation of the oral microbiome as a biomarker for early detection of human oral carcinomas. In: 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). Kansas City, MO: IEEE; 2017. doi:10.1109/bibm.2017.8217979
22. Lavanya L, Chandra J. Oral cancer analysis using machine learning techniques. Int J Eng Res Technol 2019;12(5):596-601.
23. Adeoye J, Tan JY, Choi SW, Thomson P. Prediction models applying machine learning to oral cavity cancer outcomes: a systematic review. Int J Med Inform 2021;154:104557. doi:10.1016/j.ijmedinf.2021.104557
24. Mahmood H, Shaban M, Indave BI, Santos-Silva AR, Rajpoot N, Khurram SA. Use of artificial intelligence in diagnosis of head and neck precancerous and cancerous lesions: a systematic review. Oral Oncol 2020;110:104885. doi:10.1016/j.oraloncology.2020.104885
25. Abu Eid R, Landini G. Quantification of the global and local complexity of the epithelial-connective tissue interface of normal, dysplastic, and neoplastic oral mucosae using digital imaging. Pathol Res Pract 2003;199(7):475-82. doi:10.1078/0344-0338-00448
26. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155(8):529-36. doi:10.7326/0003-4819-155-8-201110180-00009
27. Stone JC, Glass K, Munn Z, Tugwell P, Doi SA. Comparison of bias adjustment methods in meta-analysis suggests that quality effects modeling may have less limitations than other approaches. J Clin Epidemiol 2020;117:36-45. doi:10.1016/j.jclinepi.2019.09.010
28. Furuya-Kanamori L, Xu C, Hasan SS, Doi SA. Quality versus risk-of-bias assessment in clinical research. J Clin Epidemiol 2021;129:172-5. doi:10.1016/j.jclinepi.2020.09.044
29. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics 1994;50(4):1088-101.
30. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315(7109):629-34. doi:10.1136/bmj.315.7109.629
31. Alanazi AA, Khayyat MM, Khayyat MM, Elamin Elnaim BM, Abdel-Khalek S. Intelligent deep learning enabled oral squamous cell carcinoma detection and classification using biomedical images. Comput Intell Neurosci 2022;2022:7643967. doi:10.1155/2022/7643967
32. Diamant A, Chatterjee A, Vallières M, Shenouda G, Seuntjens J. Deep learning in head & neck cancer outcome prediction. Sci Rep 2019;9(1):2764. doi:10.1038/s41598-019-39206-1
33. Kann BH, Likitlersuang J, Bontempi D, Ye Z, Aneja S, Bakst R, et al. Screening for extranodal extension in HPV-associated oropharyngeal carcinoma: evaluation of a CT-based deep learning algorithm in patient data from a multicentre, randomised de-escalation trial. Lancet Digit Health 2023;5(6):e360-9. doi:10.1016/s2589-7500(23)00046-8
34. Suh CH, Lee KH, Choi YJ, Chung SR, Baek JH, Lee JH, et al. Oropharyngeal squamous cell carcinoma: radiomic machine-learning classifiers from multiparametric MR images for determination of HPV infection status. Sci Rep 2020;10(1):17525. doi:10.1038/s41598-020-74479-x
35. Chu CS, Lee NP, Adeoye J, Thomson P, Choi SW. Machine learning and treatment outcome prediction for oral cancer. J Oral Pathol Med 2020;49(10):977-85. doi:10.1111/jop.13089
36. Lin H, Chen H, Weng L, Shao J, Lin J. Automatic detection of oral cancer in smartphone-based images using deep learning for early diagnosis. J Biomed Opt 2021;26(8):086007. doi:10.1117/1.Jbo.26.8.086007
37. Shan J, Jiang R, Chen X, Zhong Y, Zhang W, Xie L, et al. Machine learning predicts lymph node metastasis in early-stage oral tongue squamous cell carcinoma. J Oral Maxillofac Surg 2020;78(12):2208-18. doi:10.1016/j.joms.2020.06.015
38. Warin K, Limprasert W, Suebnukarn S, Jinaporntham S, Jantana P, Vicharueang S. AI-based analysis of oral lesions using novel deep convolutional neural networks for early detection of oral cancer. PLoS One 2022;17(8): e0273508. doi:10.1371/journal.pone.0273508
39. Marsden M, Weyers BW, Bec J, Sun T, Gandour-Edwards RF, Birkeland AC, et al. Intraoperative margin assessment in oral and oropharyngeal cancer using label-free fluorescence lifetime imaging and machine learning. IEEE Trans Biomed Eng 2021;68(3):857-68. doi:10.1109/tbme.2020.3010480
40. Deif MA, Attar H, Amer A, Elhaty IA, Khosravi MR, Solyman AA. Diagnosis of oral squamous cell carcinoma using deep neural networks and binary particle swarm optimization on histopathological images: an AIoMT approach. Comput Intell Neurosci 2022;2022:6364102. doi:10.1155/2022/6364102
41. Fujima N, Andreu-Arasa VC, Meibom SK, Mercier GA, Salama AR, Truong MT, et al. Prediction of the treatment outcome using machine learning with FDG-PET image-based multiparametric approach in patients with oral cavity squamous cell carcinoma. Clin Radiol 2021;76(9):711.e1-7. doi:10.1016/j.crad.2021.03.017
42. Cheng NM, Yao J, Cai J, Ye X, Zhao S, Zhao K, et al. Deep learning for fully automated prediction of overall survival in patients with oropharyngeal cancer using FDG-PET imaging. Clin Cancer Res 2021;27(14):3948-59. doi:10.1158/1078-0432.Ccr-20-4935
43. Fujima N, Andreu-Arasa VC, Meibom SK, Mercier GA, Truong MT, Sakai O. Prediction of the human papillomavirus status in patients with oropharyngeal squamous cell carcinoma by FDG-PET imaging dataset using deep learning analysis: a hypothesis-generating study. Eur J Radiol 2020;126:108936. doi:10.1016/j.ejrad.2020.108936
44. Fujima N, Andreu-Arasa VC, Meibom SK, Mercier GA, Truong MT, Hirata K, et al. Prediction of the local treatment outcome in patients with oropharyngeal squamous cell carcinoma using deep learning analysis of pretreatment FDG-PET images. BMC Cancer 2021;21(1):900. doi:10.1186/s12885-021-08599-6
45. Bos P, van den Brekel MW, Gouw ZA, Al-Mamgani A, Taghavi M, Waktola S, et al. Improved outcome prediction of oropharyngeal cancer by combining clinical and MRI features in machine learning models. Eur J Radiol 2021;139:109701.

doi:10.1016/j.ejrad.2021.109701
46. Praveen Birur N, Song B, Sunny SP, Keerthi G, Mendonca P, Mukhia N, et al. Field validation of deep learning-based point-of-care device for early detection of oral malignant and potentially malignant disorders. Sci Rep 2022;12(1):14283. doi:10.1038/s41598-022-18249-x
47. Mohan R, Rama A, Raja RK, Shaik MR, Khan M, Shaik B, et al. OralNet: fused optimal deep features framework for oral squamous cell carcinoma detection. Biomolecules 2023;13(7):1090. doi:10.3390/biom13071090
48. Alabi RO, Almangush A, Elmusrati M, Leivo I, Mäkitie AA. An interpretable machine learning prognostic system for risk stratification in oropharyngeal cancer. Int J Med Inform 2022;168:104896. doi:10.1016/j.ijmedinf.2022.104896
49. Alabi RO, Elmusrati M, Sawazaki-Calone I, Kowalski LP, Haglund C, Coletta RD, et al. Comparison of supervised machine learning classification techniques in prediction of locoregional recurrences in early oral tongue cancer. Int J Med Inform 2020;136:104068. doi:10.1016/j.ijmedinf.2019.104068
50. Alabi RO, Elmusrati M, Sawazaki-Calone I, Kowalski LP, Haglund C, Coletta RD, et al. Machine learning application for prediction of locoregional recurrences in early oral tongue cancer: a web-based prognostic tool. Virchows Arch 2019;475(4):489-97. doi:10.1007/s00428-019-02642-5
51. Alabi RO, Mäkitie AA, Pirinen M, Elmusrati M, Leivo I, Almangush A. Comparison of nomogram with machine learning techniques for prediction of overall survival in patients with tongue cancer. Int J Med Inform 2021;145:104313. doi:10.1016/j.ijmedinf.2020.104313
52. Yang SY, Li SH, Liu JL, Sun XQ, Cen YY, Ren RY, et al. Histopathology-based diagnosis of oral squamous cell carcinoma using deep learning. J Dent Res 2022;101(11):1321-7. doi:10.1177/00220345221089858
53. Ranjbar S, Ning S, Zwart CM, Wood CP, Weindling SM, Wu T, et al. Computed tomography-based texture analysis to determine human papillomavirus status of oropharyngeal squamous cell carcinoma. J Comput Assist Tomogr 2018;42(2):299-305. doi:10.1097/rct.0000000000000682
54. Bae S, Choi YS, Sohn B, Ahn SS, Lee SK, Yang J, et al. Squamous cell carcinoma and lymphoma of the oropharynx: differentiation using a radiomics approach. Yonsei Med J 2020;61(10):895-900. doi:10.3349/ymj.2020.61.10.895
55. Fati SM, Senan EM, Javed Y. Early diagnosis of oral squamous cell carcinoma based on histopathological images using deep and hybrid learning approaches. Diagnostics (Basel) 2022;12(8):1899. doi:10.3390/diagnostics12081899
56. Sunny S, Baby A, James BL, Balaji D, Aparna NV, Rana MH, et al. A smart tele-cytology point-of-care platform for oral cancer screening. PLoS One 2019;14(11): e0224885. doi:10.1371/journal.pone.0224885
57. 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. doi:10.1038/s41598-023-29204-9
58. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2015;136(5): E359-86. doi:10.1002/ijc.29210
59. Allen C, Duffy S, Teknos T, Islam M, Chen Z, Albert PS, et al. Nuclear factor-κB-related serum factors as longitudinal biomarkers of response and survival in advanced oropharyngeal carcinoma. Clin Cancer Res 2007;13(11):3182-90. doi:10.1158/1078-0432.Ccr-06-3047
60. Leibowitz MS, Andrade Filho PA, Ferrone S, Ferris RL. Deficiency of activated STAT1 in head and neck cancer cells mediates TAP1-dependent escape from cytotoxic T lymphocytes. Cancer Immunol Immunother 2011;60(4):525-35. doi:10.1007/s00262-010-0961-7
61. Kuss I, Hathaway B, Ferris RL, Gooding W, Whiteside TL. Decreased absolute counts of T lymphocyte subsets and their relation to disease in squamous cell carcinoma of the head and neck. Clin Cancer Res 2004;10(11):3755-62. doi:10.1158/1078-0432.Ccr-04-0054
62. Whiteside TL. Immunobiology of head and neck cancer. Cancer Metastasis Rev 2005;24(1):95-105. doi:10.1007/s10555-005-5050-6
63. Mahmood H, Shaban M, Rajpoot N, Khurram SA. Artificial Intelligence-based methods in head and neck cancer diagnosis: an overview. Br J Cancer 2021;124(12):1934-40. doi:10.1038/s41416-021-01386-x
64. Belfiore MP, Nardone V, D'Onofrio I, Salvia AA, D'Ippolito E, Gallo L, et al. Diffusion-weighted imaging and apparent diffusion coefficient mapping of head and neck lymph node metastasis: a systematic review. Explor Target Antitumor Ther 2022;3(6):734-45. doi:10.37349/etat.2022.00110
65. Tran BX, Vu GT, Ha GH, Vuong QH, Ho MT, Vuong TT, et al. Global evolution of research in artificial intelligence in health and medicine: a bibliometric study. J Clin Med 2019;8(3):360. doi:10.3390/jcm8030360
66. Howard FM, Kochanny S, Koshy M, Spiotto M, Pearson AT. Machine learning-guided adjuvant treatment of head and neck cancer. JAMA Netw Open 2020;3(11): e2025881. doi:10.1001/jamanetworkopen.2020.25881
67. Tsutsumi K, Soltanzadeh-Zarandi S, Khosravi P, Goshtasbi K, Djalilian HR, Abouzari M. Machine learning in the management of lateral skull base tumors: a systematic review. J Otorhinolaryngol Hear Balanc Med 2022;3(4):7. doi:10.3390/ohbm3040007
68. Rokhshad R, Keyhan SO, Yousefi P. Artificial intelligence applications and ethical challenges in oral and maxillo-facial cosmetic surgery: a narrative review. Maxillofac Plast Reconstr Surg 2023;45(1):14. doi:10.1186/s40902-023-00382-w
69. Giannitto C, Carnicelli G, Lusi S, Ammirabile A, Casiraghi E, De Virgilio A, et al. The use of artificial intelligence in head and neck cancers: a multidisciplinary survey. J Pers Med 2024;14(4):341. doi:10.3390/jpm14040341