REVIEW

Editorial Process: Submission:12/28/2023 Acceptance:08/20/2024

Diagnostic Accuracy of Artificial Intelligence Compared to Biopsy in Detecting Early Oral Squamous Cell Carcinoma: A Systematic Review and Meta Analysis

Mehak Malhotra[1], Amar Kumar Shaw[2]*, Smita R Priyadarshini[1], Samruddhi Swapnil Metha[3], Pradyumna Kumar Sahoo[4], Arti Gachake[5,6]

Abstract

Objective: To summarize and compare the existing evidence on diagnostic accuracy of artificial intelligence (AI) models in detecting early oral squamous cell carcinoma (OSCC). **Method:** Review was performed in accordance to Preferred Reporting Items for Systematic Reviews and Meta-Analysis – Diagnostic Test Accuracy (PRISMA- DTA) checklist and the review protocol is registered under PROSPERO(CRD42023456355). PubMed, Google Scholar, EBSCOhost were searched from January 2000 to November 2023 to identify the diagnostic potential of AI based tools and models. True-positive, false-positive, true-negative, false-negative, sensitivity, specificity values were extracted or calculated if not present for each study. Quality of selected studies was evaluated based on QUADAS (Quality assessment of diagnostic accuracy studies)- 2 tool. Meta-analysis was performed in Meta-Disc 1.4 software and Review Manager 5.3 RevMan using a bivariate model parameter for the sensitivity and specificity and summary points, summary receiver operating curve (SROC), diagnostic odds ratio (DOR) confidence region, and area under curve (AUC) were calculated. **Results:** Fourteen studies were included for qualitative synthesis and for meta-analysis. Included studies had presence of low to moderate risk of bias. Pooled sensitivity and specificity of 0.43 (CI 0.18- 0.71) and 0.50 (CI 0.20- 0.80) was observed with a pooled positive likelihood ratio of (PLR) 0.86 (0.43 – 1.71) and negative likelihood ratio (NLR) of 1.04 (0.42 – 1.68) was observed with DOR of 0.78 (0.12 – 5.18) and overall accuracy (AUC) being 0.45 respectively. **Conclusion:** AI based tools has poor to moderate overall diagnostic accuracy. However, to validate our study findings further more standardized diagnostic accuracy studies should be conducted with proper reporting through QUADAS-2 tool. Thus, we can conclude AI based based tool for secondary level of prevention for early OSCC under early diagnosis and prompt treatment.

Keywords: Accuracy- artificial intelligence- diagnosis- meta-analysis- oral cancer

Asian Pac J Cancer Prev, **25 (8)**, 2593-2603

Introduction

Oral cancer is an aggressive disease characterized by a low average survival rate. Developments in treatment modalities in the domains of both oncology and surgery have only contributed to a rather limited improvement in outcome. Therefore, accurate diagnosis and prognosis prediction of cancer, especially at an early stage is important in improving survival rate [1].

The incidence rate of oral squamous cell carcinoma (OSCC) is increasing in many Asian countries due to the frequent consumption of alcohol and excessive tobacco chewing. OSCC is an important subtype of oral cancer that represents above 90% of total oral cancer cases [2]. Like all other cancer diagnosis, the histological evaluation, i.e. the study of tissue samples of affected region under the microscope, is the clinical practice to diagnose OSCC and its different grades. This method of microscopic investigation is referred as gold standard in cancer diagnosis. According to WHO, OSCC has been categorized into three groups; early, moderate and late-grade/stage depending on different histological

*[1]BDS, Bharati Vidyapeeth (Deemed to be University), Dental College and Hospital, Pune, Maharashtra, India. [2]Assistant Professor, Department of Public Health Dentistry Bharati Vidyapeeth (Deemed to be University), Dental College and Hospital, Pune, Maharashtra, India. [3]Professor, Department of Oral Medicine and Radiology, Institute of Dental sciences, Siksha O Anusandhan University, Bhubaneswar, Odisha, India. [4]Assistant Professor, Department of Oral Medicine and Radiology Bharati Vidyapeeth (Deemed to be University), Dental College and Hospital, Sangli, Maharashtra, India. [5]Professor, Department of Prosthodontics, Institute of Dental Sciences, Siksha O Anusandhan University, Bhubaneshwar, Odisha, India. [6]Assistant Professor, Bharati Vidyapeeth (Deemed to be University), Dental College and Hospital, Pune, Maharashtra, India. *For Correspondence: shawamar96@gmail.com*

parameters [3]. In conventional practice, pathologists used to investigate histopathological images of an oral mucosa under lower magnification (2× or 4× or 5× objective lens with effective magnification of 20×, 40×, and 50×) to find any abnormality associated with oral submucous fibrosis (OSF), oral epithelial dysplasia, or OSCC [4].

Several quantitative approaches have been reported for cancer diagnosis using biopsy followed by histopathological images [5] but a limited researches reported for automated identification of histological parameter to detect oral cancer using computer vision approaches [6, 7].

The AI approach was found to be beneficial in the three aspects that are essential to early diagnosis and prognosis. These are an improved accuracy of cancer susceptibility, recurrence, and survival predictions [8], which improve the survival rates through the effective clinical management of patients. Understanding the refinements of innovations like Artificial Intelligence (AI) could relieve potential clinical entanglements [9, 10]. Application of AI in the oral malignant growths can improve the current challenges in the disease diagnosis, as well as in predicting the prognosis. AI, which mimics human cognitive functions, is a forward leap in innovation, and has enamored the minds of scientists over the globe [5]. Its use in dentistry has begun recently, which has led to extraordinary accomplishments. History goes back to as early as 400 BC; Plato visualized an essential model of brain function. AI system is a framework that takes information, discovers designs, uses data to train itself, and yields results [11].

Understanding the diagnostic accuracy would help clinicians to reach correct diagnosis and choose most effective treatment. Diagnostic accuracy includes sensitivity, specificity and summary receiver operating characteristics (SROC) analysis [12].

Sensitivity and specificity explain the diagnostic ability of a test to correctly identify diseased and non-diseased respectively. They are independent of disease prevalence which refers to the probability of disease in a specific population at a given time and summary receiver operating characteristics (SROC) analysis is used to evaluate the predictive power for diagnosis [13].

There have been already few reviews published on various AI models for cancer detection in neck and head region [14, 15, 16]. However, these reviews could not provide validate research evidence due to presence of considerable amount of heterogeneity in the quantitative synthesis for measuring the overall diagnostic accuracy of AI based tools. Till date, no studies have provided a comprehensive, quantitative analysis of diagnostic potential of AI based on which diagnostic reasoning of early oral squamous cell carcinoma can be established. Therefore, the aim of this systematic review is to compare the diagnostic accuracy of various artificial intelligence models for early diagnosis of OSCC in adults through a meta- analysis. This gives an overview of the current status of AI based models in OSCC.

Materials and Methods

Protocol and Registration

The systematic review and meta-analysis protocol was registered at the international prospective register of systematic reviews (PROSPERO- CRD42023456355) and performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis – Diagnostic Test Accuracy (PRISMA- DTA) checklist [17].

Study Design

The following focused research question in the Participants (P), Index test (I), reference standard (R) and target condition (T) format was proposed “Is there a difference in the diagnostic accuracy of artificial intelligence tools (Index Test) compared to biopsy (gold standard) for the early detection of oral squamous cell carcinoma (OSCC) in adults?

Eligibility Criteria

studies were selected based on the following criteria:

Inclusion Criteria

The inclusion criteria were as follows:

(1) Study Design: In-vivo studies- Observational studies or cross-sectional studies comparing the diagnostic accuracy of artificial intelligence.

(2) Participant characteristics: patients diagnosed with oral squamous cell carcinoma aged 18 years and older

(3) Outcome measurements: Diagnostic accuracy including sensitivity, specificity, accuracy, determined using different methods irrespective of the methods of quantifying the outcomes.

(4) Articles written in English language

(5) Articles from January 2000 – November 2023 and available as free full text

Exclusion Criteria

The exclusion criteria were as follows:

(1) Non-clinical studies, in-vitro studies, and animal studies. Studies reporting about a single intervention were also excluded.

(2) Studies done on individuals less than 18 years of age.

(3) Studies not fully available in the database.

(4) Article reporting only abstracts were also excluded.

(5) Studies not reporting primary outcomes of accuracy, sensitivity, and specificity as well as where primary outcomes are not possible to calculate from the given raw data.

Search protocol and study selection

A comprehensive electronic search was performed till November 2023 for the studies published within the last 23 years (from 2000 to 2023) using the following databases: PubMed and EBSCOhost to retrieve articles in the English language. The searches in the clinical trials database, cross-referencing and grey literature were conducted using Google Scholar, Greylist, and OpenGrey. In addition to the electronic search, a hand search was also made, and reference lists of the selected articles were screened.

Search Strategy

Appropriate key words and Medical Subject Heading (MeSH) terms were selected and combined with Boolean operators like AND. The search strategy used was as follows: (artificial intelligence AND sensitivity AND specificity AND oral cancer), (histopathology AND malignant transformation AND diagnosis).

Search Strategy according to PIRT Format

	Strategy
Population	(("oral cancer diagnosis"[MeSH Terms] OR "oral cancer prediction"[All Fields] OR "malignant transformation"[All Fields]) OR ("child"[MeSH Terms] OR"child"[AllFields]))OR("automaticdiagnosis"[MeSH Terms] OR ("residual local adaptation"[All Fields]
Index test	((("artificial intelligence"[MeSH Terms] OR ("artificial neural network"[All Fields] AND "machine learning"[All Fields] AND "deep learning"[All Fields]) OR "diagnostic accuracy"[All Fields] OR "explainable AI"[All Fields]) OR ("local binary pattern"[MeSH Terms] OR ("fuzzy neural network"[All Fields] AND "therapy"[All Fields]) OR "fuzzy regression"[All Fields])) OR ("hybrid method")[MeSH Terms]
Reference standard	(("histopathology "[All Fields] OR "oral biopsy"[MeSH Terms] OR ("artificial neural network"[All Fields] AND "treatment"[All Fields]) OR "oral tissue"[All Fields] OR ("smartphone-based learning"[All Fields] AND "anaesthesia"[All Fields]) OR "transfer leaning"[All Fields]))
Target condition	((("oral cancer prediction"[MeSH Terms] OR "head and neck cancer"[All Fields]) OR ("oral cancer"[MeSH Terms] OR ("dysplasia"[All Fields] AND "oral potentially malignant disorder" [All Fields]) OR "oral squamous cell carcinoma"[All Fields] OR ("comparative study"[All Fields] AND "randomized controlled trial"[All Fields] AND "clinical study"[All Fields]) OR "prospective study"[All Fields]))

The search and screening, according to the previously established protocol were conducted by two review authors. A two-phase selection of articles was conducted. In phase one, two reviewers reviewed titles and abstracts of all articles. Articles that did not meet inclusion criteria were excluded. In phase-two, selected full articles were independently reviewed and screened by same reviewers. Any disagreement was resolved by discussion. When mutual agreement between two reviewers was not reached, a third reviewer was involved to make final decision. The final selection was based on consensus among all three authors.

Data extraction

For all included studies, following descriptive study details were extracted by two independent reviewing authors (and) using pilot-tested customized data extraction forms: authors, year of study, country, study design, sample size, AI model (Index test), sensitivitry (%), specificity (%) and conclusion. Quantitative data of sensitivity and specificity were compiled from each study and using these quantitative data, values like true positive, true negative, false positive and false negatives were calculated manually for the studies using the below formulas where the data was not provided by authors. The corresponding authors were contacted via email where further information was needed [13].

a) False positive = (1-specificity) x (1- diseased cases/ total sample)

b) True negative = specificity x (1- diseased cases/ total sample)

c) True positive = sensitivity x diseased cases/ total sample

d) False negative = (1- sensitivity) x diseased cases/ total sample

Assessment of methodological quality

The methodological quality or the risk of bias was evaluated using Quality Assessment for Diagnostic Accuracy Studies -2 (QUADAS-2) tool [18]. The QUADAS-2 is a revised tool developed to assess quality of diagnostic studies through its four domains: patient selection, index test, reference standard, flow and timing of participants. Each domain had signalling questions with options of “Yes”, “No” or “Unclear”. The overall risk of bias was assessed as high: if answered ‘No’ to any question, Low: if answered ‘Yes’ to all questions and Unclear: if answered ‘Unclear’ to all questions or accompanied by any ‘Yes’. Risk of bias summary and applicability concern was graphically plotted using Review Manager (RevMan) software version 5.3.

Statistical analysis and data synthesis

Raw data was used to calculate sensitivity and specificity for each biomarker with their estimation method. For overall accuracy, we calculated pooled sensitivity, pooled specificity with 95% confidence interval, area under summary receiver operating characteristic. (Interpretation of AUC values were as follows: value above 80% were considered as excellent, between 70% and 80% as good, between 60% and 69% as fair and below 60% as poor outcomes for a diagnostic test [19]. To assess the impact of heterogeneity, Higgins I2 test was used. This test represents the proportion of variability due to heterogeneity rather than due to sampling error [20]. According to I2 test statistic the heterogeneity could be low (I2 <50%) or high (I2 >50%). Subgroup analysis was also carried out. Results were presented graphically as coupled forest plot for each salivary biomarker with their estimation method using Meta-Disc 1.4 software.

Additional analysis

Additional analysis was performed with positive likelihood ratio (PLR) and negative likelihood ratio (NLR) using DerSimonian-Laird’s estimator considering random effect model. Positive likelihood ratio (PLR) in range of 2-5, 5-10 and >10 represents small, moderate and large increase in probability of disease when test is positive while Negative likelihood ratio (NLR) in range of 0.2-0.5, 0.2-0.1 and <0.1 represents small, moderate and large decrease in probability of disease when test is negative [21].

Results

Study Selection

After duplicates removal, reference list of included studies was screened. Of which 121 studies were excluded. After this full text articles were assessed for eligibility and articles that did not meet inclusion criteria were excluded. Fourteen studies fulfilled eligibility criteria and were included in qualitative synthesis and in meta – analysis. A flowchart of identification, inclusion and exclusion of studies is shown in Figure 1.

Study Characteristics

A summary of descriptive characteristics of all included studies is shown in Table 1. Data was evaluated from fourteen studies from an aggregate of 7047 specimens for which the diagnostic accuracy values of various AI based models (artificial neural network (ANN), coherence optical tomography (COT), decision tree (DT), support vector machine (SVM), Gaussian Mixture Model (GMM), K-Nearest Neighbour (KNN), Local Residual Adaptation Network (LRAN), Radial Basis Probabilistic Neural Network (RBPNN) was compared with biopsy followed by histopathological investigations. All the studies had cross-sectional comparative study design with four studies [7, 8, 2, 10] were conducted in China, three studies [9, 1, 11] were conducted in India and one study [22] in Malaysia, Baik et al. [4] in Canada, Heidari et al. [6] in Finland, Chu et al. [3] in Hong Kong, Amin et al. [5] in Brazil, Fati et al. [23] in Saudi Arabia and Jubair et al. [24] in Jordan. The overall sensitivity and specificity was observed in the range of 41.98 – 100% and 45.5 – 100% respectively. It was concluded that AI tool has great potential in predicting and diagnosing disease outcome and AI can improve the quality and reach of oral cancer screening and early detection.

Risk of Bias within Studies

Patient selection was considered as high risk of bias for Amin et al. [5], which was mainly due to method of patient enrollment, nature of study design and implementing inappropriate exclusion. Remaining all studies, reported low risk of bias with respect to patient selection domain.

The index test was considered to be at low risk of bias among all the included studies. High risk of bias is usually reported with respect to index test domain when there is insufficient details reported as to whether results of index test was interpreted without prior knowledge of reference standard results, lack of pre-specification of a test-positive threshold and statement of conflict of interest.

Similarly, one study [5] reported high risk of bias for the reference standard; while a low risk was seen among other included studies and flow and timing domain was considered at low risk in all studies. The risk of bias and applicability concern summary and graph is depicted in Figure 2, 3.

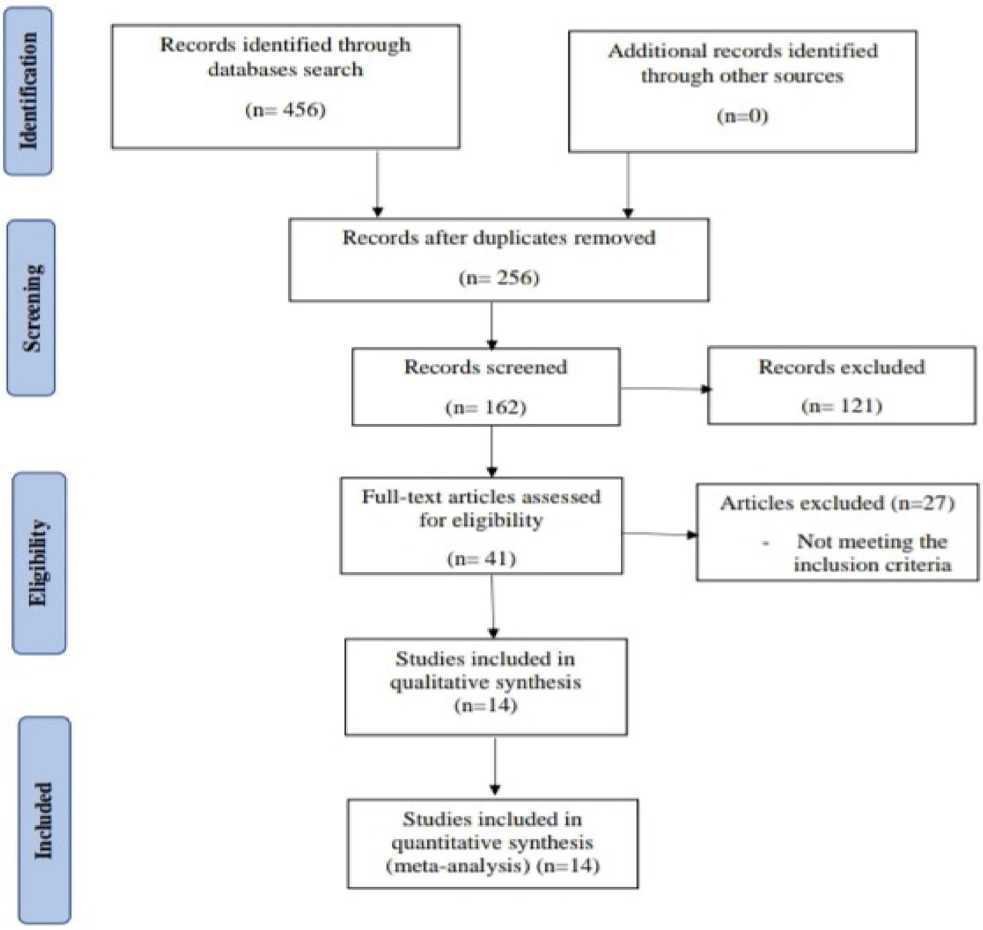

Figure 1. PRISMA Flow Diagram

Table 1. Showing Descriptive Characteristics of Included Studies

Author, years of study	Country	Study design	Sample size	AI model (Index test)	Sensitivity (%)	Specificity (%)	Conclusion
Nayak et al. [9]	India	Comparative study	84	PCA & ANN	100	96	the methods are very attractive for real time applications
Rosma et al. [22]	Malaysia	Comparative study	84	fuzzy neural network model and fuzzy regression model	59.9	45.5	both fuzzy regression and fuzzy neural network models provide good alternative to human expert prediction in predicting oral cancer susceptibility.
Krishnan et al. [1]	India	Cross-sectional study	42	Sugeno Fuzzy, GMM, K-NN, RBPNN	94.5	98.8	AI tool has great potential in predicting and diagnosing disease outcome
Baik et al. [4]	Canada	Cross-sectional study	28	Forest based algorithms	80.6	79.3	a crucial asset in the implementation of high-resolution image analysis in routine clinical pathology practice to identify lesions
Heidari et al. [6]	Finland	Cross-sectional study	10	OCT based algorithms	89	100	This image processing algorithm and mobile imaging system could provide a useful screening and triage tool for basic level field screeners where specialist expertise and facilities are not available
Chu et al. [3]	Hong kong	Comparative study	467	DT, SVM, KNN models	41.98	84.12	Machine learning helps clinicians in assessing the and predicting disease outcome
Fu et al. [7]	China	Cross-sectional study	1469	automated deep learning algorithm using cascaded convolutional neural networks	94.9	88.7	deep learning methods may offer opportunities for automatically identifying OCSCC patients with the performance matching or even beyond that of skilled human experts.
Amin et al. [5]	Brazil	Cross-sectional study	934 OSCC images	Deep learning (DL) model	95.16	95	AI can improve the quality and reach of oral cancer screening and early detection
Lin et al. [8]	China	Cross-sectional study	65	Deep learning (DL) network	83	96.6	The smartphone-based imaging with deep learning method has good potential for primary oral cancer diagnosis
James et al. [11]	India	Comparative study	75	Optical coherence tomography (OCT) and ANN	95	93	potential clinical application of device in screening and surveillance of oral cancer.
Warin et al. [2]	China	Comparative study	350	DenseNet 121 and R-CNN model	98.75	100	DenseNet 121 proved to offer acceptable potential for detection of cancerous lesions in oral photographic images
Fati et al. [23]	Saudi Arabia	Comparative study	2698 histopathological images	CNN models and SVM algorithms	99.5	99.61	the tremendous potential of artificial intelligence techniques to diagnose OSCC and increase cure rates among patients
Jubair et al. [24]	Jordan	Comparative study	716	Deep CNNs	86.7	84.5	AI can improve the quality and reach of oral cancer screening and early detection
Yuan et al. [10]	China	Comparative study	25	LRAN models	91.66	92.58	LRAN model has excellent capability to solve the non-invasive oral cancer screening task.

ANN, artificial neural network; COT, coherence optical tomography; DT, decision tree; SVM, support vector machine; GMM, Gaussian Mixture Model; KNN, K-Nearest Neighbour; LRAN, Local Residual Adaptation Network; RBPNN, Radial Basis Probabilistic Neural Network

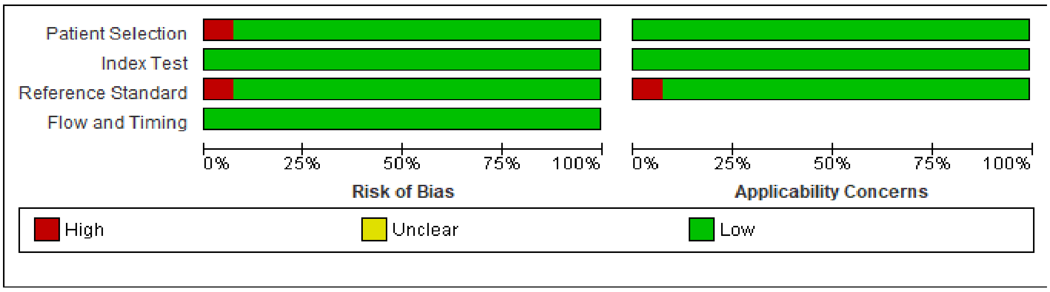

Figure 2. The Risk of Bias and Applicability Concern Summary and Graph

Synthesis of result

This meta-analysis was conducted for evaluating the overall diagnostic accuracy of AI in patients with OSCC. Summary statistics measure was calculated in terms of pooled sensitivity, specificity, positive and negative likelihood ratio (PLR & NLR), diagnostic odd's ratio (DOR) and area under the curve (AUC).

As shown in Figure 4, data was evaluated from fourteen studies [5, 8, 9, 11, 23, 22, 3, 2, 1, 24, 7, 6, 10, 4] investigating the overall diagnostic accuracy. The pooled sensitivity was 0.43 (CI 0.18- 0.71) and pooled specificity was 0.50 (CI 0.20- 0.80) with I2 being 0%.

As shown in Figure 5. the area under the curve (AUC) was plotted with sensitivity and 1-specificity and standard error. An overall accuracy of (AUC) 0.45 was seen for AI indicating that the AI had a fair to poor efficacy in diagnosing the condition.

Additional analysis

Likelihood ratio was estimated which signifies the ability of the index test to predict the test results (positive / negative) when the disease condition in actual is present or absent. As shown in Figure 6, pooled positive likelihood ratio (PLR) 0.86 (0.43 – 1.71) and negative likelihood ratio (NLR) 1.04 (0.42 – 1.68) was estimated. Pooled PLR suggested that test result is associated with absence of disease when the disease is present while pooled NLR suggested that the test result is associated with presence of disease when the disease is absent.

As shown in Figure 7 the pooled Diagnostic Odds Ratio (DOR) is 0.78 (0.12 – 5.18) suggesting that overall ability of index test in correctly diagnosing the target condition is moderate.

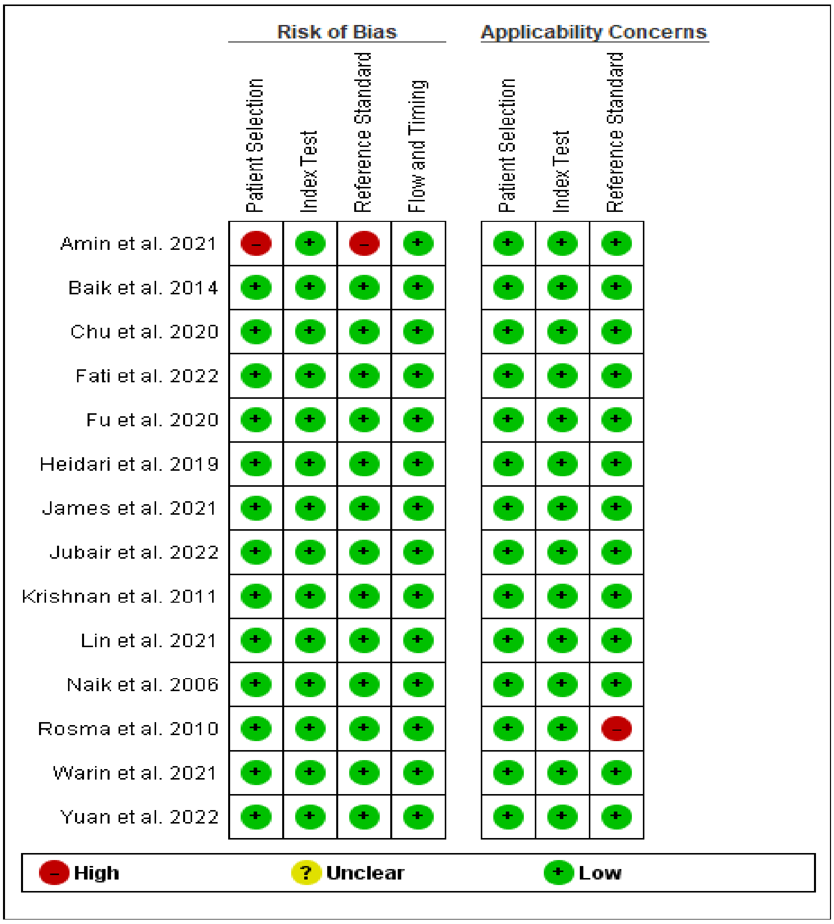

Figure 3. Risk of Bias and Applicability Concerns Graph: Review Authors' Judgements about Each Domain Presented as Percentages Across Included Studies

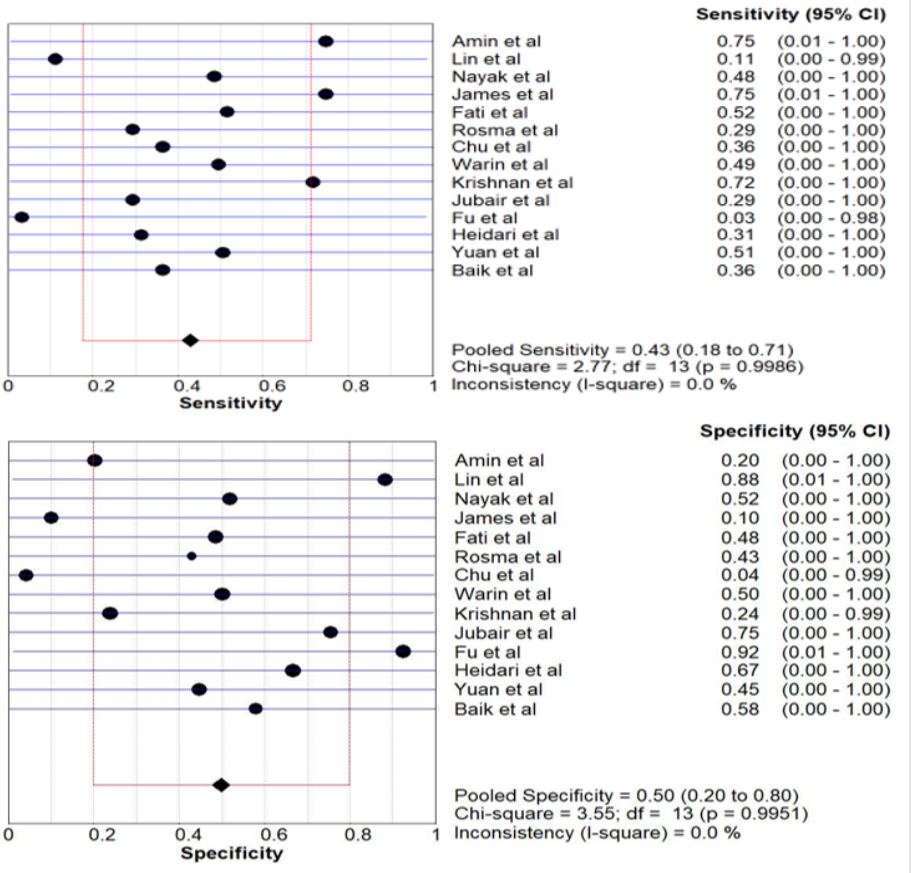

Figure 4. Pooled Sensitivity and Specificity for AI

Discussion

Early detection and regular surveillance of suspicious oral lesions are critical for decreasing mortality rate of OSCC. The current gold standard for OSCC diagnosis is the histopathological assessment of biopsied oral tissue [5], although other imaging techniques can be used sometimes to complement the detection and staging of the lesions. Nonetheless, a substantial number of flaws are associated with the current modalities. First, besides the invasive nature of biopsies, they are subjected to sampling errors, which may lead to misdiagnosis [8]. Second, difficulty in locating the region due to the non-uniform appearance causes most OSCC to be detected

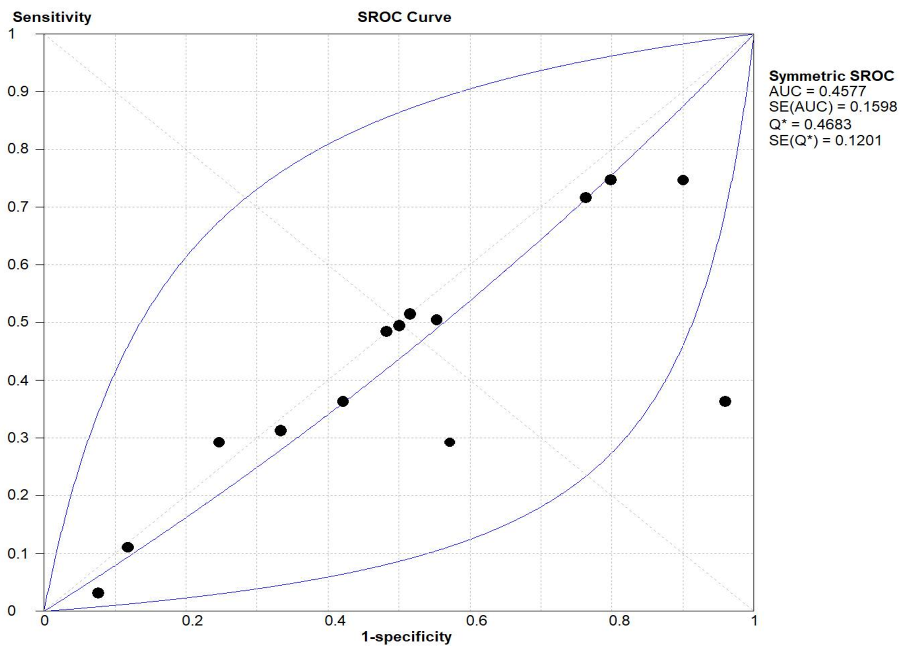

Figure 5. Overall Accuracy through Area under the Curve (AUC) with Summary Receiver Operating Characteristics (SROC) Curve was Plotted for AI Models

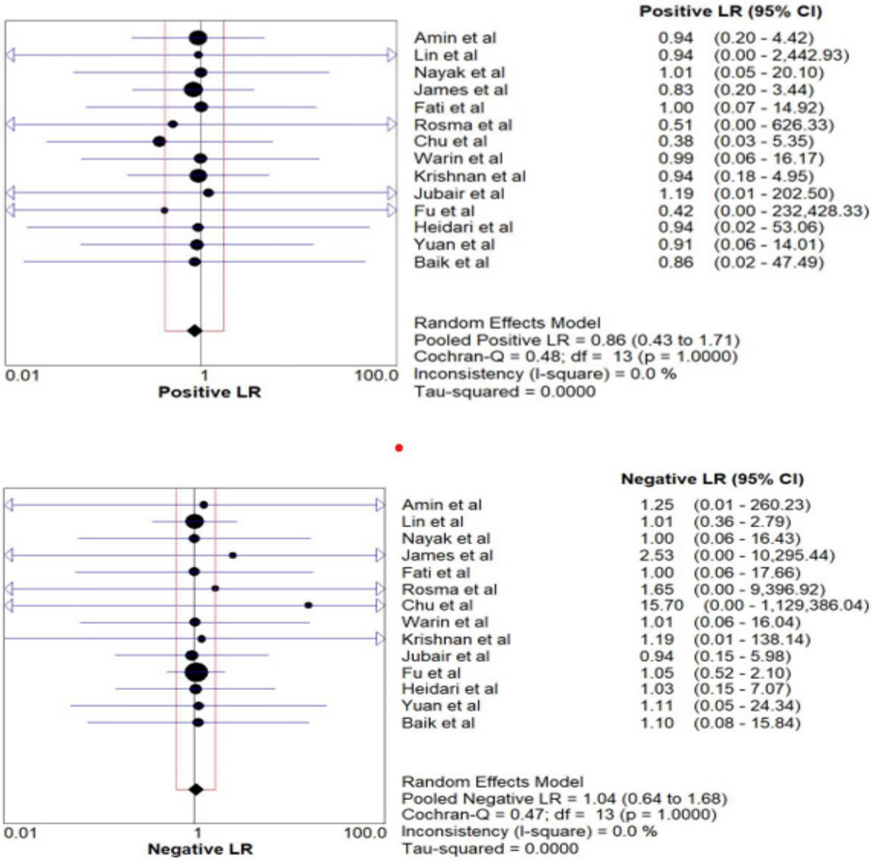

Figure 6. Showing Positive and Negative Likelihood Ratio’s for AI based Models

when cancer has already advanced to late stages [4]. Third, intratumor heterogeneity in OSCC often requires evaluation by qualified pathologists, and despite the potential of identifying suspicious lesions, the shortage of trained professionals and healthcare resources limits access and makes the OSCC burden fall on the developing nations [3].

Since early diagnosis has been correlated with better outcomes and survival, making a quick and efficient diagnosis is, therefore, a major step in the course of patient management [14]. Multiple adjunctive diagnostic aids reported in the literature have provided some potential. Their accuracy has further improved with the advancements in machine learning. With the rapid development of computer algorithms, AI has been increasingly used to enhance the early diagnosis of OSSC using different modalities.

Mahmood et al. [14], conducted a systematic review

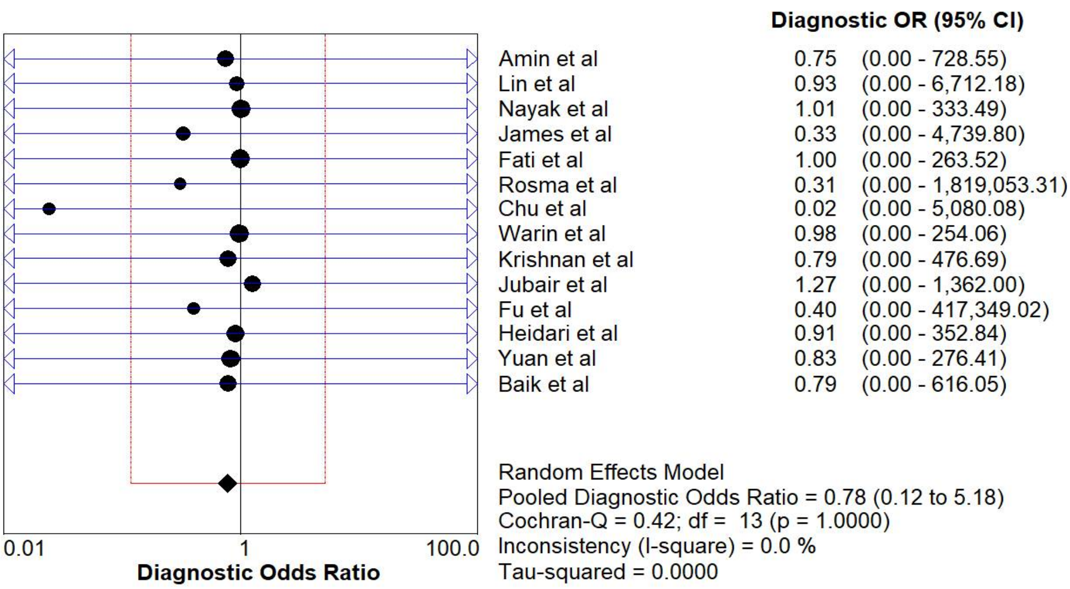

Figure 7. Showing Diagnostic Odds Ratio (DOR’s) for AI based Models

to analyse and describe the application and diagnostic accuracy of Artificial Intelligence (AI) methods used for detection and grading of potentially malignant (pre-cancerous) and cancerous head and neck lesions. Electronic databases were searched from October 2009 - April 2020. 11 studies were included in final review with an accuracy between 79-100%. The review provided early evidence to support the potential application of supervised machine learning methods as a diagnostic aid for some oral potentially malignant and malignant lesions.

Alabi et al. [25], conducted a systematic review of diagnostic and prognostic application of AI and machine learning in oral squamous cell carcinoma (OSCC) and also highlights some of the limitations and concerns of clinicians towards the implementation of machine learning-based models for daily clinical practice. Electronic databases were searched till February 2020. 41 original studies fulfilled the eligibility criteria and were included in the review. Majority of these studies used the support vector machine (SVM) and artificial neural network (ANN) algorithms. A specificity ranging from 0.57 to 1.00, sensitivity from 0.70 to 1.00, and an accuracy from 63.4 % to 100.0 % was observed among the studies. It was concluded that these models reported to show promising performances for diagnostic and prognostic analyses in studies of oral cancer and these models should be developed to further enhance explainability, interpretability, and externally validated for generalizability in order to be safely integrated into daily clinical practices.

Khanagar et al. [15], conducted a systematic review on the application and performance of AI in diagnosis and predicting the occurrence of Oral cancer (OC). Databases were searched from January 2000 to March 2021 yielding 16 studies. It was found that AI can accurately predict the occurrence of OC, as compared to conventional methods. The precision and accuracy of AI in diagnosis as well as predicting the occurrence are higher than the current, existing clinical strategies. Elmakaty et al. [26], carried out a systematic review and meta-analysis to evaluate the accuracy of artificial intelligence (AI)-assisted technologies in detecting OSCC. Six databases like PubMed, Embase, Scopus, Cochrane Library, ProQuest, and Web of Science up to 15 Mar 2022 yielding 16 studies with twelve different AI models. The sensitivity, specificity, positive and negative likelihood ratios as well as the pooled diagnostic odds ratio were 92.0 % (95 % confidence interval [17] 86.7–95.4 %), 91.9 % (95 % CI 86.5–95.3 %), 11.4 (95 % CI 6.74–19.2), 0.087 (95 % CI 0.051–0.146) and 132 (95 % CI 62.6–277), respectively. The results of study supported the capability of AI-assisted systems to detect OSCC with high accuracy, potentially aiding the histopathological examination in early diagnosis.

Khanagar et al. [16], conducted a systematic review to critically appraise the available evidence regarding the utilization of AI in the diagnosis, classification, and prediction of oral cancer (OC) using histopathological images. Databases were searched from January 2000 and January 2023. Nineteen studies were included in review reported to have an overall accuracy in a range from 89.47% to 100%, sensitivity from 97.76% to 99.26%, and specificity ranging from 92% to 99.42%. It was concluded that AI has a superior level of precision and accuracy, helping pathologists significantly improve their diagnostic outcomes and reduce the probability of errors. The aim of this systematic review and meta-analysis is to summarize existing evidence on AI based tools and to compare their accuracy in diagnosing early oral squamous cell carcinoma in adults. To the best of our knowledge, this is the first systematic review and meta-analysis which provides a comprehensive quantitative analysis of AI in early oral squamous cell carcinoma diagnosis. A total of 7047 specimens from fourteen eligible studies were included in meta-analysis. Most of the AI based tools had good diagnostic accuracy. To further evaluate their diagnostic accuracy, we calculated positive and negative likelihood ratio. Furthermore, we also conducted a diagnostic odds ratio analysis.

Fourteen studies fulfilled the eligibility criteria and were included in review, for which the diagnostic accuracy values of various AI based models (artificial neural network (ANN), coherence optical tomography (COT), decision tree (DT), support vector machine (SVM), Gaussian Mixture Model (GMM), K-Nearest Neighbour (KNN), Local Residual Adaptation Network (LRAN), Radial Basis Probabilistic Neural Network (RBPNN) was compared with biopsy followed by histopathological investigations. Result of review concluded that AI tool has great potential in predicting and diagnosing disease outcome and AI can improve the quality and reach of oral cancer screening and early detection. However, meta-analysis revealed a pooled sensitivity and specificity of 0.43 (CI 0.18- 0.71) and 0.50 (CI 0.20- 0.80) respectively with a pooled positive likelihood ratio of (PLR) 0.86 (0.43 – 1.71) and negative likelihood ratio (NLR) of 1.04 (0.42 – 1.68) was observed with DOR of 0.78 (0.12 – 5.18) and overall accuracy (AUC) being 0.45 suggesting that the overall diagnostic accuracy of AI based tools being poor to moderate in diagnosing the desired condition.

Most of included studies were at high risk of selection bias arising from use of a 'case-control' study design. In addition, patient sampling and/or recruitment into studies were insufficiently reported. Among the included studies, only three studies [5] had sufficiently reported patient selection process. All studies used biopsy/histopathological investigation as reference standard and AI tools as index test. However, insufficient detail and lack of clarity in reporting studies made it difficult to assess risk of bias. Therefore, use of STARD (Standards for Reporting of Diagnostic Accuracy Studies) [27] checklist in reporting primary studies could have facilitated the quality appraisal. Reporting guidelines for primary diagnostic studies should be followed strictly and studies should address all potential source of bias and applicability concern as indicated in QUADAS-2 tool [18].

This study is limited by overall quality of included studies. Further standardised diagnostic test accuracy studies that minimises potential sources of bias through rigorous design, conduct and reporting are needed. Future research must focus on the accuracy of current potential principal salivary biomarkers in detection of OSCC with clear and robust methodology. The adherence to the PRISMA guidelines, the thorough unrestricted literature search, utilization of reliable methodology with regard to the qualitative synthesis of data, the quality assessment of evidence with the Cochrane risk of bias tool for randomized controlled trials strengthens this systematic review. The quality assessment of all the included studies showed low-moderate risk of bias whereas overall quality was high, specifying lack of potential and inevitable sources of bias with limited variability and reporting deficiencies.

A systematic review is a transparent and repeatable procedure for identifying, selecting and critically assessing published or unpublished data to address a well-defined research question. Meta-analyses, a statistical analysis that incorporates numerical data from related studies, are frequently paired with systematic reviews. The best evidence is generally regarded as systematic reviews and meta-analyses. However, the calibre of the included studies has an impact on how strong the evidence is from a systematic review and meta-analysis. In the current systematic review, sufficient studies with a brief observation period and a known risk of bias were included. As a result, the presently available evidence is sufficient to make therapeutic recommendations in response to the current systematic review's focus question.

In conclusion, from the results of study, it was concluded that AI based tools has poor to moderate overall diagnostic accuracy. However, to validate our study findings further more standardized diagnostic accuracy studies should be conducted with proper reporting through STARD checklist and QUADAS-2 tool. Therefore, we can conclude AI based told for secondary level of prevention for early OSCC under early diagnosis and prompt treatment.

Author Contribution Statement

All authors contributed equally in this study.

Acknowledgements

None.

References

1. Krishnan MM, Venkatraghavan V, Acharya UR, Pal M, Paul RR, Min LC, et al. Automated oral cancer identification using histopathological images: A hybrid feature extraction paradigm. Micron. 2012;43(2-3):352-64. https://doi.org/10.1016/j.micron.2011.09.016.
2. Warin K, Limprasert W, Suebnukarn S, Jinaporntham S, Jantana P. Automatic classification and detection of oral cancer in photographic images using deep learning algorithms. J Oral Pathol Med. 2021;50(9):911-8. https://doi.org/10.1111/jop.13227.
3. 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. https://doi.org/10.1111/jop.13089.
4. Baik J, Ye Q, Zhang L, Poh C, Rosin M, MacAulay C, et al. Automated classification of oral premalignant lesions using image cytometry and random forests-based algorithms. Cell Oncol (Dordr). 2014;37(3):193-202. https://doi.org/10.1007/s13402-014-0172-x.
5. Amin I, Zamir H, Khan F. Histopathological image analysis for oral squamous cell carcinoma classification using concatenated deep learning models. medRxiv. 2021;15.
6. Heidari ae, sunny sp, james bl, lam tm, tran av, yu j, ramanjinappa rd, et al. Optical coherence tomography as an oral cancer screening adjunct in a low resource settings. 2018. Ieee J Sel Top Quantum Electron;25(1):1-8.
7. Fu Q, Chen Y, Li Z, Jing Q, Hu C, Liu H, et al. A deep learning algorithm for detection of oral cavity squamous cell carcinoma from photographic images: A retrospective study. EClinicalMedicine. 2020;27:100558. https://doi.org/10.1016/j.eclinm.2020.100558.
8. 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). https://doi.org/10.1117/1.Jbo.26.8.086007.
9. Nayak GS, Kamath S, Pai KM, Sarkar A, Ray S, Kurien J, et al. Principal component analysis and artificial neural network analysis of oral tissue fluorescence spectra: Classification of normal premalignant and malignant pathological conditions. Biopolymers. 2006;82(2):152-66. https://doi.org/10.1002/bip.20473.
10. Yuan W, Cheng L, Yang J, Yin B, Fan X, Yang J, et al. Noninvasive oral cancer screening based on local residual adaptation network using optical coherence tomography. Med Biol Eng Comput. 2022;60(5):1363-75. https://doi.org/10.1007/s11517-022-02535-x.
11. James BL, Sunny SP, Heidari AE, Ramanjinappa RD, Lam T, Tran AV, et al. Validation of a point-of-care optical coherence tomography device with machine learning algorithm for detection of oral potentially malignant and malignant lesions. Cancers (Basel). 2021;13(14). https://doi.org/10.3390/cancers13143583.
12. Shaw AK, Garcha V, Shetty V, Vinay V, Bhor K, Ambildhok K, et al. Diagnostic accuracy of salivary biomarkers in detecting early oral squamous cell carcinoma: A systematic review and meta-analysis. Asian Pac J Cancer Prev. 2022;23(5):1483-95. https://doi.org/10.31557/apjcp.2022.23.5.1483.
13. Shaw ak, mahajan m, varshney s, jena m, rohatgi l, bashir s, tewari s, roy j. Diagnostic accuracy of chemiluminescence for oral potentially malignant disorders: A systematic review and meta-analysis. J Clin Diagn Res. 2022;16(7).
14. 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. https://doi.org/10.1016/j.oraloncology.2020.104885.
15. Khanagar SB, Naik S, Al Kheraif AA, Vishwanathaiah S, Maganur PC, Alhazmi Y, et al. Application and performance of artificial intelligence technology in oral cancer diagnosis and prediction of prognosis: A systematic review. Diagnostics (Basel). 2021;11(6). https://doi.org/10.3390/diagnostics11061004.
16. 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). https://doi.org/10.3390/biomedicines11061612.
17. Salameh JP, Bossuyt PM, McGrath TA, Thombs BD, Hyde CJ, Macaskill P, et al. Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (prisma-dta): Explanation, elaboration, and checklist. BMJ. 2020;370:m2632. https://doi.org/10.1136/

bmj.m2632.
18. 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. https://doi.org/10.7326/0003-4819-155-8-201110180-00009.
19. Jones CM, Athanasiou T. Summary receiver operating characteristic curve analysis techniques in the evaluation of diagnostic tests. Ann Thorac Surg. 2005;79(1):16-20. https://doi.org/10.1016/j.athoracsur.2004.09.040.
20. Lijmer JG, Bossuyt PM, Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. Stat Med. 2002;21(11):1525-37. https://doi.org/10.1002/sim.1185.
21. Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. Lancet. 2005;365(9469):1500-5. https://doi.org/10.1016/s0140-6736(05)66422-7.
22. Mohd Dom R, Abdul Kareem S, Abidin B, Mazlipah S, Norzaidi M. The use of artificial intelligence to identify people at risk of oral cancer: Empirical evidence in malaysian university. Int J Sci Res. 2010;3:10-20.
23. 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). https://doi.org/10.3390/diagnostics12081899.
24. Jubair F, Al-Karadsheh O, Malamos D, Al Mahdi S, Saad Y, Hassona Y. A novel lightweight deep convolutional neural network for early detection of oral cancer. Oral Dis. 2022;28(4):1123-30. https://doi.org/10.1111/odi.13825.
25. Alabi RO, Youssef O, Pirinen M, Elmusrati M, Mäkitie AA, Leivo I, et al. Machine learning in oral squamous cell carcinoma: Current status, clinical concerns and prospects for future-a systematic review. Artif Intell Med. 2021;115:102060. https://doi.org/10.1016/j.artmed.2021.102060.
26. 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. https://doi.org/10.1016/j.critrevonc.2022.103777.
27. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. Stard 2015 guidelines for reporting diagnostic accuracy studies: Explanation and elaboration. BMJ Open. 2016;6(11):e012799. https://doi.org/10.1136/bmjopen-2016-012799.