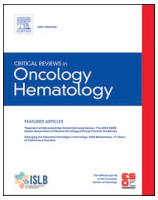
Critical Reviews in Oncology / Hematology 178 (2022) 103777

Contents lists available at ScienceDirect

Critical Reviews in Oncology / Hematology

journal homepage: www.elsevier.com/locate/critrevonc



Accuracy of artificial intelligence-assisted detection of Oral Squamous Cell Carcinoma: A systematic review and meta-analysis

Ibrahim Elmakaty [a], Mohamed Elmarasi [a], Ahmed Amarah [a], Ruba Abdo [a], Mohammed Imad Malki [b,*,1]

[a] College of Medicine, QU Health, Qatar University, Doha, Qatar
[b] Pathology Unit, Department of Basic Medical Sciences, College of Medicine, QU Health, Qatar University, Doha, Qatar

ARTICLE INFO

Keywords:
Diagnostic accuracy
Artificial intelligence
Oral cancers
Squamous Cell Carcinoma
Meta-analysis

ABSTRACT

Oral Squamous Cell Carcinoma (OSCC) is an aggressive tumor with a poor prognosis. Accurate and timely diagnosis is therefore essential for reducing the burden of advanced disease and improving outcomes. In this meta-analysis, we evaluated the accuracy of artificial intelligence (AI)-assisted technologies in detecting OSCC. We included studies that validated any diagnostic modality that used AI to detect OSCC. A search was performed in six databases: PubMed, Embase, Scopus, Cochrane Library, ProQuest, and Web of Science up to 15 Mar 2022. The Quality Assessment Tool for Diagnostic Accuracy Studies was used to evaluate the included studies' quality, while the Split Component Synthesis method was utilized to quantitatively synthesize the pooled diagnostic efficacy estimates. We considered 16 out of the 566 yielded studies, which included twelve different AI models with a total of 6606 samples. The summary sensitivity, summary specificity, positive and negative likelihood ratios as well as the pooled diagnostic odds ratio were 92.0 % (95 % confidence interval [CI] 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. Our findings support the capability of AI-assisted systems to detect OSCC with high accuracy, potentially aiding the histopathological examination in early diagnosis, yet more prospective studies are needed to justify their use in the real population.

1. Introduction

Oral Cancer is one of the most prevalent cancers that imposes a considerable disease burden across the world. The Global Cancer Observatory (GLOBCAN) most recent estimates for oral cavity and lip cancers indicate that there were 377,713 new cases and 177,757 deaths in 2020, ranking as the sixteenth most common cancer worldwide (Sung et al., 2021). Oral Squamous Cell Carcinoma (OSCC) constitutes the vast majority of oral cancer cases, accounting for more than 90 % overall (Coletta et al., 2020). Over the past years, OSCC incidence has been increasing in developing countries, among females, and alarmingly among adults younger than the age of 45 years old (Coletta et al., 2020). This can be explained, in part, by the increasing consumption of tobacco and alcohol in some countries and among females, despite that, marked geographical and environmental risk factor variations still exist (Coletta et al., 2020).

OSCC generally develops from precursor lesions termed Oral Potential Malignant Disorders (OPMDs) mainly represented by oral lichen planus, leukoplakia, and erythroplakia (Lin et al., 2021). Still, nearly 40 % of patients first present with advanced OSCC stage IV, for which combined chemoradiotherapy is required (Cheng et al., 2020). This makes OSCC associated with a poor survival rate and prognosis despite the advancements made in the understanding of the involved biological processes and the therapeutic options (Coletta et al., 2020). Hence, recognizing OPMDs and diagnosing OSCC at an early stage is of paramount importance to developing effective preventive strategies and improving clinical management.

To date, OSCC diagnosis is made by clinical history, conventional oral examination followed by an incisional biopsy of the suspected tissue. This approach, however, has limitations in that the lesions are clinically heterogeneous, and their interpretation is subjective depending on the examiner's experience, while the biopsy is invasive with

* Correspondence to: College of Medicine, QU Health, Qatar University, P. O. Box: 2713, Doha, Qatar.
E-mail addresses: ie1703006@qu.edu.qa (I. Elmakaty), me1803937@qu.edu.qa (M. Elmarasi), aa1704483@qu.edu.qa (A. Amarah), ra1701813@qu.edu.qa (R. Abdo), momalki@qu.edu.qa (M.I. Malki).
[1] 0000-0002-6801-2126.

https://doi.org/10.1016/j.critrevonc.2022.103777
Received 23 June 2022; Received in revised form 21 July 2022; Accepted 1 August 2022
Available online 2 August 2022

potential complications (Lin et al., 2021). Recently, the machine learning approach has exhibited a comparative advantage over the traditional diagnostic methods. Numerous studies have reported their improved accuracy of cancer susceptibility, and outcome prediction (Alabi et al., 2021). Although the feasibility of an automated approach can reduce errors and promote informed decisions, there is little consensus on whether it can be reliably applied in the actual healthcare setting. Thus, this study aims to evaluate the diagnostic accuracy of artificial intelligence (AI)-assisted detection of OSCC in-vivo in comparison with gold-standard histopathology.

2. Methods

2.1. Protocol and registration

This systematic review and meta-analysis was reported following The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline (see PRISMA checklist in Supplementary material, Table S1) (Moher et al., 2009). The review protocol was registered with the international prospective register of systematic reviews (PROSPERO) online database (PROSPERO Identifier: CRD42022333367).

2.2. Search strategy

We developed our search strategy in the PubMed database using Medical Subject Headings (MeSH) terms and keywords derived from four key concepts. These are AI, diagnosis, oral, and cancer. There were no language or date restrictions imposed on the search. The Polyglot translator (Clark et al., 2020) was used to transfer the developed search strategy to Embase, Scopus, Cochrane Library, ProQuest, and Web of science. The full strategy for each database is available in the Supplementary material. All the yielded studies were then transferred to EndNote X7, where duplicates were identified and consequently eliminated. To ensure that relevant studies were not overlooked, the reference lists of all eligible articles were manually reviewed.

2.3. Eligibility criteria

We implemented predefined broad eligibility criteria and included both prospective and retrospective studies that validated the diagnostic performance of any AI modality in detecting OSCC in-vivo compared to the gold standard biopsy histopathology. Studies were included if they fulfilled the following criteria: reported the values of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), or values were possibly calculated from the sensitivity (Se) and specificity (Sp). In addition to studies that had AI models that can distinguish between OSSC and any other tissue type in a binary manner, whether normal, precancerous, or cancerous. Studies that did not meet these inclusion criteria were excluded. We excluded articles that aimed to investigate the accuracy in classifying, segmenting, delineating, predicting, and differentiating metastases or recurrence. In case a study assessed one of these parameters along with the diagnostic accuracy, only data of the latter was used. We further excluded studies that had animal models, low sample size, foreign language, or others including systematic reviews, meta-analyses, clinical trials, and randomized controlled trials. Articles that used the same datasets were considered duplicates, and only the article with the larger sample size was eventually included.

2.4. Study selection and screening

The remaining studies were uploaded to the Rayyan platform for further screening (Ouzzani et al., 2016). Two reviewers independently screened titles and abstracts, and any disagreement was addressed by consensus among them. The full texts of papers deemed eligible were then obtained and independently double-screened, with any inconsistencies handled through discussion with the whole team.

2.5. Data extraction

We extracted data from each eligible study on the first author, publication date, study design, participants' demographics (age, sex), country, AI index test, reference test, TP, FP, TN, FN, validation, and training datasets, and OSSC prevalence. The diagnostic accuracy of multiple AI models of a single study was extracted separately. If a dataset for a model was validated multiple times, the highest Se and Sp were selected. This was done by two reviewers independently, with a third-party opinion if needed. All data were summarized and compiled into an online Microsoft Excel spreadsheet that was accessible to all the authors.

2.6. Quality of studies

The methodological quality of selected diagnostic accuracy studies was assessed using the Quality Assessment Tool for Diagnostic Accuracy Studies (QUADAS-2) (Whiting et al., 2011), which was done independently by two reviewers with group discussion when necessary. We deemed the tool's relevant questions to relate to bias safeguards, and instead of grading the risk of bias as a judgment, we tallied the number of safeguards implemented in each study. Thus, if implemented ("yes/-low bias"), a numeric value of 1 was assigned, and if not implemented ("no/high bias/unclear"), a numeric value of 0 was assigned. This was done so that the quality assessment can be used in a bias-adjusted meta-analysis model with quantitative utility (Furuya-Kanamori et al., 2021b; Stone et al., 2020). In QUADAS-2, the judgment question assigned in each domain (patient selection, index test, reference standard, flow, and timing) was removed to decrease subjectivity in the bias-adjusted synthesis. As a result of this approach, fourteen comprehensive bias safeguards were produced. The quality assessment scoring sheet is provided in the Supplementary material (Table S2).

2.7. Data analysis

For bias-adjusted synthesis, the Quality Effect (QE) model was used (Doi et al., 2015; Doi and Thalib, 2008), with the results of the quality assessment represented as relative rankings (Stone et al., 2021). By dispersing study weights by quality rank, the QE model controls for methodological quality-related heterogeneity, hence bias adjusting the synthesis (Doi et al., 2015; Doi and Thalib, 2008). Additionally, the Se, Sp, positive likelihood ratio (pLR), negative likelihood ratio (nLR), and diagnostic odds ratio (DOR) with their 95 % confidence intervals (95 % CI) were calculated using the split component synthesis method (SCS) for meta-analysis of diagnostic studies (Furuya-Kanamori et al., 2021a). Summary receiver operating characteristic (sROC) plots were also created using the SCS method (Furuya-Kanamori et al., 2021a). The I2 statistic was used to examine heterogeneity (Higgins et al., 2003), while Doi plots were used to analyze publication bias and putative small study effects (Furuya-Kanamori et al., 2018). The LFK index was used to quantify Doi plots symmetry (Furuya-Kanamori et al., 2018). After installing the diagma module (Furuya-Kanamori et al., 2021a), Stata version 16 was utilized for all analyzes, graphs, and plots.

2.8. Subgroup analysis

To assess the performance of each index test on its own and compare it with the overall diagnostic accuracy for all modalities, the extracted datasets were divided into separate groups if a test had greater than or equal to five different sets. Accordingly, three groups were created for Fluorescence Spectroscopy, Raman Spectroscopy, and Oral Photographic Images.

3. Results

3.1. Study selection

Fig. 1 depicts the PRISMA flow diagram that demonstrates the process of study selection. Our exhaustive database search initially yielded 566 articles. Of which, 129 were eliminated through EndNote, and the remaining 437 were screened for title and abstract. Following the title and abstract screening, 371 publications were excluded, leaving only 66 for full-text screening. The full texts of these 66 records were retrieved and reviewed for eligibility. For various reasons summarized in Fig. 1, 52 articles were excluded. Hand-searching of reference lists of the 14 studies left resulted in the addition of two more studies, bringing the total number of the included studies in our meta-analysis to 16 (Bhowmik et al., 2022; Cals et al., 2016; Fu et al., 2020; Heintzelman et al., 2000; Kamath and Mahato, 2007; Lin et al., 2021; Majumder et al., 2005; Mohamed et al., 2021; Nayak et al., 2006; Song et al., 2020; Sunny et al., 2019; Wang et al., 2020; Warin et al., 2021; Welikala et al., 2020; Yu et al., 2019; Zhou et al., 2021). Further details on the citation of the excluded records with full reasoning are provided in the Supplementary material, Table S3.

3.2. Study and index test characteristics

Table 1 shows the extracted datasets of the included articles. In brief, they were published between 2005 and 2021, with the majority coming from China (n = 6) and India (n = 5). Ten of the sixteen studies were retrospective in nature, where past samples/images were evaluated, while the remaining six articles were prospective. Two studies focused on spectra samples from tongue OSCC. Image Analysis (n = 5) was the most commonly used test for diagnosing OSCC, followed by Fluorescence Spectroscopy (n = 3) and Raman Spectroscopy (n = 3). The other tests in our selected studies were Mass Spectrometry, Tele-cytology, scoring based on a non-invasive examination, Oral Microbiota Gene Testing, and a Breath Analyzer. There were 12 different AI models used among the included datasets. The most common of these include CNN

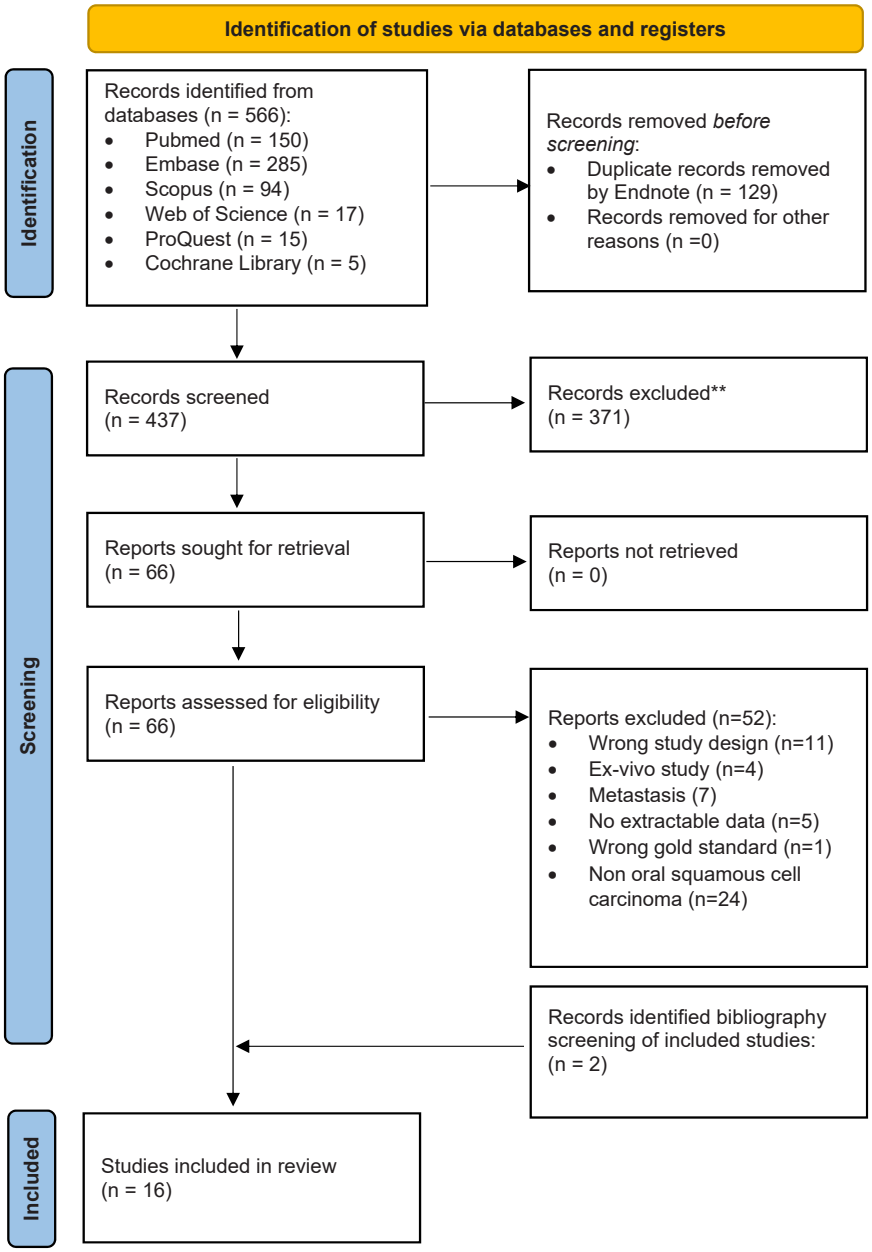


Fig. 1. PRISMA flow-chart for the systematic review and meta-analysis.

Table 1
Characteristics and extracted data from included studies.

Authors (year)	Country	Study type	Cancer Site	Machine learning model	Index test	Samples/ Images (n)	Training (n)	Validation (n)	Patients (n)	TP	FP	FN	TN
(Nayak et al., 2006)	India	Retrospective	Oral cavity	ANN	Autofluorescence spectroscopy	143	60	83	6	35	0	2	46
(Nayak et al., 2006)	India	Retrospective	Oral cavity	PCA	Autofluorescence spectroscopy	143	60	83	6	36	0	1	46
(Wang et al., 2020)	China	Retrospective	Oral cavity	Random Forest	Scoring system* (baseline Model)	266	159	107	266	27	6	7	67
(Wang et al., 2020)	China	Retrospective	Oral cavity	Random Forest	Scoring system** (personalized Model)	266	159	107	266	28	6	6	67
(Song et al., 2020)	China	Prospective	Oral cavity	Lasso regression	Saliva Mass spectrometry	373	193	180	373	46	3	14	117
(Sunny et al., 2019)	India	Prospective	Oral cavity	SVM	Tele-cytology system	11,981	532	30	60	14	2	1	13
(Sunny et al., 2019)	India	Prospective	Oral cavity	Random Forest	Tele-cytology system	11,981	532	30	60	14	3	1	12
(Sunny et al., 2019)	India	Prospective	Oral cavity	Logistic regression	Tele-cytology system	11,981	532	30	60	12	2	3	13
(Sunny et al., 2019)	India	Prospective	Oral cavity	LDA	Tele-cytology system	11,981	532	30	60	12	2	3	13
(Sunny et al., 2019)	India	Prospective	Oral cavity	KNN	Tele-cytology system	11,981	532	30	60	11	4	4	12
(Fu et al., 2020) (Internal validation)	China	Retrospective	Oral cavity	Cascaded CNN	Photographs	7244	5775	401	NM	170	25	9	197
(Fu et al., 2020) (external validation)	China	Retrospective	Oral cavity	Cascaded CNN	Photographs	7244	5775	402	NM	138	48	16	200
(Fu et al., 2020) (clinical validation)	China	Retrospective	Oral cavity	Cascaded CNN	Photographs	7244	5775	666	NM	249	25	25	367
(Yu et al., 2019)	China	Prospective	Tongue	CNN	Fiber optic Raman spectroscopy	1440	1152	288	12	143	8	1	136
(Yu et al., 2019)	China	Prospective	Tongue	SVM	Fiber optic Raman spectroscopy	1440	1152	288	12	139	8	5	136
(Yu et al., 2019)	China	Prospective	Tongue	LDA	Fiber optic Raman spectroscopy	1440	1152	288	12	141	17	3	127
(Cals et al., 2016)	Netherlands	Prospective	Tongue	PCA	Raman spectroscopy	1087	720	367	21	54	68	0	245
(Zhou et al., 2021)	China	Prospective	Oral cavity	Random Forest	Oral microbiota rDNA sequencing	146	75	18	93	659	97	0	756
(Warin et al., 2021)	Thailand	Retrospective	Oral cavity	CNN	Oral photographs	700	490	70	NM	69	0	1	70
(Welikala et al., 2020)	Malaysia	Retrospective	Oral cavity	CNN	Oral photographs	2155	1744	207	1085	11	6	13	174
(Heintzelman et al., 2000)	USA	Retrospective	Oral cavity	Multivariate Discriminant Analysis	Fluorescence spectroscopy	343	62	281	76	277	0	0	4
(Mohamed et al., 2021)	Sudan	Retrospective	Oral cavity	ANN	Portable breath analyzer	84	62	27	84	8	4	2	13
(Majumder et al., 2005)	India	Prospective	Oral cavity	RVM	Autofluorescence spectroscopy	325	119	206	29	37	4	4	73
(Majumder et al., 2005)	India	Prospective	Oral cavity	SVM	Autofluorescence spectroscopy	325	119	206	29	38	4	3	73
(Lin et al., 2021)	China	Retrospective	Oral cavity	CNN based HRNet	Oral photographs	7994	7539	455	722	424	3	1	27
(Kamath and Mahato, 2007)	India	Retrospective	Oral cavity	KNN	Fiber optic Raman spectroscopy	143	60	83	6	40	0	3	40
(Bhowmik et al., 2022)	India	Retrospective	Oral cavity	Ensemble model	Blood perfusion imager	440	183	257	61	221	0	2	34

Abbreviations: TP, true positives; FP, false positives; FN, false negatives; TN, true negatives; ANN, artificial neural networks; PCA, principal component analysis; SVM, support vector machine; LDA, linear discriminant analysis; KNN, k-nearest neighbor; CNN, convolutional neural network; RVM, relevance vector machine; HRNet, high-resolution network; rDNA, ribosomal deoxyribonucleic Acid.

* Score based on Visually enhanced lesion scope and toluidine blue staining.

** Score based on Visually enhanced lesion scope and toluidine blue staining in addition to personal information and features.

(n = 7), Random Forest (n = 4), and SVM (n = 3).

3.3. Quality assessment

Overall, the number of safeguards implemented in each study ranged from 8 to 12 out of 14, with an average of 10.4 safeguards per study across the four QUADAS-2 tool domains previously mentioned. In the selection domain, the majority of papers (n = 7) utilized 3/4 safeguards. In the flow and timing domain, the highest scores were roughly distributed between 12 studies, 3/4 (n = 6) and 4/4 (n = 6). In the index test and reference standard domains, most articles scored 2/3 and 3/3, respectively. Fig. 2 depicts the safeguards applied per domain for each of the included studies, and the Supplementary material includes a table containing the answers to all quality assessment questions, Table S4.

3.4. Main analysis results

The 27 datasets included a total of 6606 samples with an OSCC prevalence of 48.2 %. The summary Se, summary Sp, pLR, nLR, and pooled DOR for overall performance of AI-assisted diagnosis of OSCC were 92.0 % (95% CI 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. For this overall group, there was a significant heterogeneity with an I^2 value of 75.9 %, however, this heterogeneity is expected as this group contains different index tests that are likely to differ in their performance. Regarding the publication bias, there was minor positive asymmetry in the Doi plot with an LKF index value of 1.89, suggesting that some of the smaller studies had slightly larger DOR than the bigger studies. This means better diagnostic accuracy results were to some extent more likely to be published. The Doi plot is available in the Supplementary material (Fig. S1), as well as the results for the rest of the analyzes (Table S5). Fig. 3 represents the sROC for all datasets (n = 27), with the circles representing the sROC and the solid square representing the summary Se and Sp intersection point. The upper and lower square boundaries represent the Se confidence limits, while the left and right boundaries represent the 1-Sp confidence limits. The area under the curve (AUC) was 92.0 % (95% CI 88.8–94.3 %).

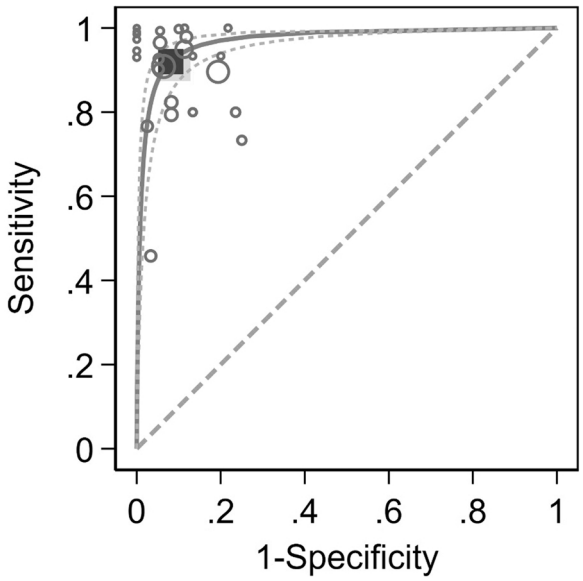


Fig. 3. Summary receiver operating characteristic curve of artificial intelligence-assisted systems for the diagnosis of OSCC.

In terms of the tests, Raman spectrometry had the best Se among all test groups with 97.8 % (n = 5 studies, samples = 1314, 95 % CI 96.1–98.8 %) followed by Fluorescence Spectroscopy which achieved a Se of 93.2 % (n = 5 studies, samples = 683, 95% CI 85.6–96.9 %). In contrast, Fluorescence Spectroscopy had the highest Sp among all groups with 96.5 % (95 % CI 92.0–98.5 %) making it the most accurate test to confirm the diagnosis of OSCC, followed by Photography then Raman Spectrometry.

Heterogeneity assessment in both Fluorescence Spectroscopy and Raman Spectrometry were insignificant with I^2 values of 18.8 % and 0 % respectively, indicating the consistency of the findings in these two

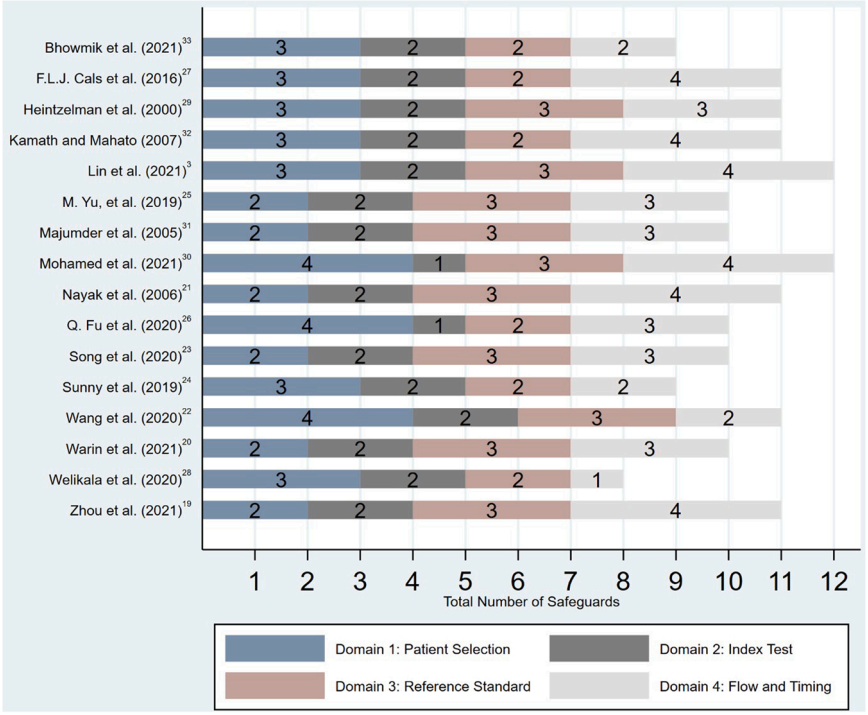


Fig. 2. Quality assessment using QUADAS-2 for included studies.

groups. The use of oral images in diagnosing OSCC achieved good performance reaching a Se of 90.4 % (n = 7 studies, samples = 2525, I^2 = 85.6 %, 95% CI 78.8–96.0 %) and great Sp with a value of 93.2 % (95 % CI 83.8–97.3 %). Fig. 4 demonstrates a bar chart that compares these groups with the overall performance of all tests.

4. Discussion

4.1. Principal findings

With the rapid development of computer algorithms, AI has been increasingly used to enhance the early diagnosis of OSSC using different modalities. To the best of our knowledge, this is the first meta-analysis that evaluated the performance of AI-assisted OSCC diagnosis. In this systematic review and meta-analysis, we combined the diagnostic accuracy of different AI-assisted models and demonstrated that they have high accuracy in differentiating healthy oral tissue from OSCC with an overall DOR of 132 (95 % CI 62.6–277). Three diagnostic tests, namely Fluorescence Spectroscopy, Raman Spectroscopy, and Photography had many studies validating them and therefore had distinct groups showing their performance. It is also worth noting that the other tests that were not put in a separate group due to the limited number of articles reported high diagnostic accuracy as well. One study used AI in gene sequencing of oral tissue and reported a model's diagnostic accuracy of 95.70 % and a prediction of 100 % in OSCC (Zhou et al., 2021). Another study recorded an accuracy of 86.7 % using conductive polymer spray mass spectrometry with the machine learning aid (Alabi et al., 2021). Tele-cytology with AI assistance also exhibited great potential in detecting OSCC (Sunny et al., 2019).

The integration of AI in oral cancer diagnostic systems has been reported to improve the accuracy of detection. A study reported a 30 % improvement when AI is combined with tele-cytology (Sunny et al., 2019). Similarly, multiple other studies compared the results of AI-assisted diagnoses with human evaluation and showed remarkable advancement when AI use (Fu et al., 2020).

4.2. Other uses of AI in OSCC

Among the recent advancements in the field, AI has been shown to have a multitude of applications other than detecting OSCC lesions (Alabi et al., 2021). One study investigated an AI-assisted model in classifying OSCC using morphological and textural features and achieved an accuracy of 99.78 % (Rahman et al., 2020a). Alternatively, in OSCC resection surgeries, AI-assisted methods have also demonstrated excellent results in the segmentation of tumor tissue (Trajanovski et al., 2021). Another crucial use of AI is its ability to predict lymph node metastasis as OSCC was found to have a frequent rate of metastasis to the cervical lymph nodes varying from 20 % to 50 % (Shan et al., 2020). Additionally, there are reports on the potential of AI technologies in predicting the 5-year overall survival rate, as well as the molecular features and prognosis of head and neck carcinoma by analyzing gene expression (Alkhadar et al., 2021; Chen et al., 2021; Zhao et al., 2020). The applications have not just been used to support diagnosis and prognosis but are also being steadily incorporated into treatment planning. The findings of one study implied that machine learning methods can be used to identify individuals with high risk who would benefit from chemotherapy and radiation (Howard et al., 2020).

We believe that all of these abovementioned uses highlight that machine learning approaches offer an extremely promising future in assisting clinicians with precision medicine and minimizing the current diagnostic limitations. This will potentially improve patient-specific therapies and support efficient, effective, and dynamic administration of hospital resources.

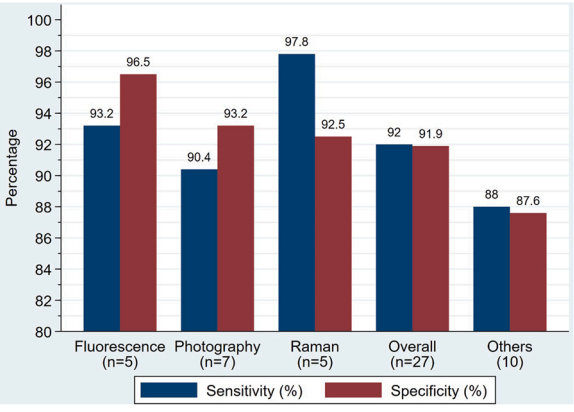


Fig. 4. Summary sensitivity and specificity for OSCC AI-assisted diagnostic modalities.

4.3. Issues in the current practice

The current gold standard for OSCC diagnosis is the histopathological assessment of biopsied oral tissue (Shah and Gil, 2009), although other imaging techniques can be used sometimes to complement the detection and staging of the lesions (Sarrión Pérez et al., 2015). 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 (Lin et al., 2021). Second, difficulty in locating the region due to the non-uniform appearance causes most OSCC to be detected when cancer has already advanced to late stages (Yakob et al., 2014). 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 (Lin et al., 2021).

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 (Carreras-Torras and Gay-Escoda, 2015). Multiple adjunctive diagnostic aids reported in the literature have provided some potential (Omar, 2015). Their accuracy has further improved with the advancements in machine learning. These include, but are not limited to, oral microbiome gene expression readers (Zhou et al., 2021), oral images through smartphones (Fu et al., 2020; Warin et al., 2021; Welikala et al., 2020), breath analyzers (Mohamed et al., 2021), and blood perfusion imagers (Bhowmik et al., 2022).

4.4. Limitations

This systematic review and meta-analysis have certain shortcomings. As previously indicated in the results section, there was a moderate degree of publication bias towards better diagnostic accuracy from the smaller studies. Furthermore, given the data limitations, we were unable to perform subgroup analysis for several diagnostic tests or evaluate their performance with a larger sample size and greater confidence. It is also important to note that one study found AI models' diagnostic accuracy to be comparable to that of experienced professionals and greater than that of medical and non-medical students (Fu et al., 2020). However, due to the inadequate number of such studies, assessing these comparisons was deemed unfeasible. There have also been very few prospective studies that incorporated AI's diagnostic accuracy in clinical settings. As a result, our results may not accurately represent AI performance in real-world patients.

Despite AI being hailed as a promising tool to revolutionize medicine, numerous ethical concerns are raised by its deployment. These include data privacy and confidentiality, informed consent, and algorithmic biases (Gerke et al., 2020). This emphasizes the need for a multidisciplinary approach to build ethical and legal guidelines to ensure a successful implementation of an AI system that people can

trust.

4.5. Policy implications and future research

Safety and Efficacy are two major elements that must be considered when adopting AI-assisted diagnostic tests. Thus, it is suggested to verify the models with datasets apart from the ones used for their training (Alabi et al., 2021). Another important aspect for machine learning models to give long-term benefits is to improve healthcare organizations' data infrastructure (Chilamkurthy et al., 2018). This will help in building more reliable models using heterogeneous and aggregated data from several sources (Chilamkurthy et al., 2018). A possible way to achieve this is by encouraging producers to publish testing samples in world bank databases that are accessible for researchers to use in creating AI-automated tools (Rahman et al., 2020b). As the majority of the studies retrieved by our literature search in this paper were retrospective, we believe that this gap must be addressed by carrying out more randomized prospective clinical studies for better validation. Nevertheless, professionals need to gain sufficient knowledge of the models' performance before execution. A validated method that might help is the decision curve analysis which demonstrates the net benefit of these models (Vickers et al., 2008). Finally, we would also recommend publishers include raw data on the TP, FP, TN, and FN values. This will assist enormously in strengthening the transparency in developing and validating multiple algorithms, and even comparing them to that of experts.

Although many OPMLs and OSCCs are discovered during routine oral exams, there is no screening program currently in use. AI can provide adequate screening for people at risk and can serve as a bridge for clinicians to become acquainted with AI applications in healthcare (Fu et al., 2020). Furthermore, AI performance in detecting OSCC has the potential to reduce human burden and time. It can even be involved in more complex decisions that often cause disagreements among pathologists primarily in determining the margin and stage of cancer tissues. Thus, AI can help to overcome the poor reproducibility and the variety of current grading and staging results among pathologists, leading to much better clinical outcomes for patients (Trajanovski et al., 2021; Welikala et al., 2020). Besides, it is able to predict metastasis and provide an accurate prognosis for better patient management, which humans are unable to do and often introduce bias (Alkhadar et al., 2021; Chen et al., 2021; Shan et al., 2020; Zhao et al., 2020). Addressing the aforementioned issues and the ethical challenges offer great potential for AI to penetrate deeper into healthcare delivery and cover population-based needs in the foreseeable future. We believe that advanced tasks in which AI can potentially play a leading role in the future of healthcare systems.

5. Conclusion

OSCC carries a poor prognosis, thus time is of the essence in reducing its burden. Unlike the current gold standard of histopathological classification, AI-assisted systems are quick, non-invasive, and have demonstrated remarkable performance in detecting OSCC. AI can also be used to help with OSCC resection, segmentation, metastasis prediction, and treatment selection. While AI-based tests demonstrated excellent accuracy in OSCC detection, providing a great opportunity to aid in the histopathological examination, they still have a long way to go before being deployed in place of the gold standard in actual medical practice. We advocate for the use of AI in oral images for screening purposes, which can serve as a bridge to familiarize clinicians with AI applications in healthcare. Additional efforts to optimize clinical workflow integration and to conduct prospective evaluation of AI-based tools in clinical settings remain important future directions.

CRediT authorship contribution statement

MIM conceived of the study idea and designed the methodology, **IE** and **ME** performed the literature search. **MIM, IE, ME, AA,** and **RA** performed the screening, data abstraction, risk of bias assessments, and Data curation. **MIM** and **IE** performed the data analysis. **MIM, IE,** and **ME** interpreted the data. All authors drafted the manuscript. **MIM** revised and supervised the drafted paper. All authors approved the final version of the manuscript.

Ethical approval

Not applicable. All the work was developed using published data.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest Disclosures

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. This paper has not been posted to any preprint repository.

Data availability statement

The data used for analysis in this work is available upon reasonable request from the corresponding author.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.critrevonc.2022.103777.

References

Alabi, R.O., Youssef, O., Pirinen, M., Elmusrati, M., Mäkitie, A.A., Leivo, I., Almangush, A., 2021. Machine learning in oral squamous cell carcinoma: current status, clinical concerns and prospects for future–a systematic review. Artif. Intell. Med. 115, 102060.

Alkhadar, H., Macluskey, M., White, S., Ellis, I., Gardner, A., 2021. Comparison of machine learning algorithms for the prediction of five-year survival in oral squamous cell carcinoma. J. Oral Pathol. Med. 50 (4), 378–384.

Bhowmik, A., Ghosh, B., Pal, M., Paul, R.R., Chatterjee, J., Chakraborty, S., 2022. Portable, handheld, and affordable blood perfusion imager for screening of subsurface cancer in resource-limited settings. Proc. Natl. Acad. Sci. USA 119 (2).

Cals, F.L., Koljenović, S., Hardillo, J.A., Baatenburg de Jong, R.J., Bakker Schut, T.C., Puppels, G.J., 2016. Development and validation of Raman spectroscopic classification models to discriminate tongue squamous cell carcinoma from non-tumorous tissue. Oral Oncol. 60, 41–47.

Carreras-Torras, C., Gay-Escoda, C., 2015. Techniques for early diagnosis of oral squamous cell carcinoma: systematic review. Med. Oral Patol. Oral cirugia Bucal 20 (3), e305–e315.

Chen, L., Zeng, H., Zhang, M., Luo, Y., Ma, X., 2021. Histopathological image and gene expression pattern analysis for predicting molecular features and prognosis of head and neck squamous cell carcinoma. Cancer Med. 10 (13), 4615–4628.

Cheng, H.C., Yang, C.C., Kao, S.Y., Wu, T.Y., Wu, C.H., 2020. Evaluation of factors associated with the risk stratification of survivorship for stage IV squamous cell carcinoma of the oral cavity: a 10-year retrospective study. J. Chin. Med. Assoc.: JCMA 83 (5), 491–499.

Chilamkurthy, S., Ghosh, R., Tanamala, S., Biviji, M., Campeau, N.G., Venugopal, V.K., Mahajan, V., Rao, P., Warier, P., 2018. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. Lancet 392 (10162), 2388–2396.

Clark, J.M., Sanders, S., Carter, M., Honeyman, D., Cleo, G., Auld, Y., Booth, D., Condron, P., Dalais, C., Bateup, S., Linthwaite, B., May, N., Munn, J., Ramsay, L., Rickett, K., Rutter, C., Smith, A., Sondergeld, P., Wallin, M., Jones, M., Beller, E., 2020. Improving the translation of search strategies using the Polyglot Search Translator: a randomized controlled trial. J. Med. Libr. Assoc.: JMLA 108 (2), 195–207.

Coletta, R.D., Yeudall, W.A., Salo, T., 2020. Grand challenges in oral cancers. Front. Oral Health 1, 3.

Doi, S.A., Thalib, L., 2008. A quality-effects model for meta-analysis. Epidemiology 19 (1), 94–100.
Doi, S.A., Barendregt, J.J., Khan, S., Thalib, L., Williams, G.M., 2015. Advances in the meta-analysis of heterogeneous clinical trials II: the quality effects model. Contemp. Clin. Trials 45 (Pt A), 123–129.
Fu, Q., Chen, Y., Li, Z., Jing, Q., Hu, C., Liu, H., Bao, J., Hong, Y., Shi, T., Li, K., Zou, H., Song, Y., Wang, H., Wang, X., Wang, Y., Liu, J., Liu, H., Chen, S., Chen, R., Zhang, M., Zhao, J., Xiang, J., Liu, B., Jia, J., Wu, H., Zhao, Y., Wan, L., Xiong, X., 2020. A deep learning algorithm for detection of oral cavity squamous cell carcinoma from photographic images: a retrospective study. EClinicalMedicine 27, 100558.
Furuya-Kanamori, L., Barendregt, J.J., Doi, S.A.R., 2018. A new improved graphical and quantitative method for detecting bias in meta-analysis. Int. J. Evid.-Based Healthc. 16 (4), 195–203.
Furuya-Kanamori, L., Kostoulas, P., Doi, S.A.R., 2021a. A new method for synthesizing test accuracy data outperformed the bivariate method. J. Clin. Epidemiol. 132, 51–58.
Furuya-Kanamori, L., Xu, C., Hasan, S.S., Doi, S.A., 2021b. Quality versus risk-of-bias assessment in clinical research. J. Clin. Epidemiol. 129, 172–175.
Gerke, S., Minssen, T., Cohen, G., 2020. Chapter 12 – ethical and legal challenges of artificial intelligence-driven healthcare. In: Bohr, A., Memarzadeh, K. (Eds.), Artificial Intelligence in Healthcare. Academic Press, pp. 295–336.
Heintzelman, D.L., Utzinger, U., Fuchs, H., Zuluaga, A., Gossage, K., Gillenwater, A.M., Jacob, R., Kemp, B., Richards-Kortum, R.R., 2000. Optimal excitation wavelengths for in vivo detection of oral neoplasia using fluorescence spectroscopy. Photochem. Photobiol. 72 (1), 103–113.
Higgins, J.P., Thompson, S.G., Deeks, J.J., Altman, D.G., 2003. Measuring inconsistency in meta-analyses. BMJ (Clin. Res. Ed.) 327 (7414), 557–560.
Howard, F.M., Kochanny, S., Koshy, M., Spiotto, M., Pearson, A.T., 2020. Machine learning-guided adjuvant treatment of head and neck cancer. JAMA Netw. Open 3 (11), e2025881.
Kamath, S.D., Mahato, K.K., 2007. Optical pathology using oral tissue fluorescence spectra: classification by principal component analysis and k-means nearest neighbor analysis. J. Biomed. Opt. 12 (1).
Lin, H., Chen, H., Weng, L., Shao, J., Lin, J., 2021. Automatic detection of oral cancer in smartphone-based images using deep learning for early diagnosis. J. Biomed. Opt. 26, 8.
Majumder, S.K., Ghosh, N., Gupta, P.K., 2005. Relevance vector machine for optical diagnosis of cancer. Lasers Surg. Med. 36 (4), 323–333.
Mohamed, N., van de Goor, R., El-Sheikh, M., Elrayah, O., Osman, T., Nginamau, E.S., Johannessen, A.C., Suleiman, A., Costea, D.E., Kross, K.W., 2021. Feasibility of a portable electronic nose for detection of oral squamous cell carcinoma in sudan. Healthcare 9, 5.
Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 6 (7), e1000097.
Nayak, G.S., Kamath, S., Pai, K.M., Sarkar, A., Ray, S., Kurien, J., D'Almeida, L., Krishnanand, B.R., Santhosh, C., Kartha, V.B., Mahato, K.K., 2006. Principal component analysis and artificial neural network analysis of oral tissue fluorescence spectra: classification of normal premalignant and malignant pathological conditions. Biopolymers 82 (2), 152–166.
Omar, E., 2015. Future imaging alternatives: the clinical non-invasive modalities in diagnosis of oral squamous cell carcinoma (OSCC). Open Dent. J. 9, 311–318.
Ouzzani, M., Hammady, H., Fedorowicz, Z., Elmagarmid, A., 2016. Rayyan-a web and mobile app for systematic reviews. Syst. Rev. 5 (1), 210.
Rahman, T.Y., Mahanta, L.B., Choudhury, H., Das, A.K., Sarma, J.D., 2020a. Study of morphological and textural features for classification of oral squamous cell carcinoma by traditional machine learning techniques. Cancer Rep. 3 (6), e1293.
Rahman, T.Y., Mahanta, L.B., Das, A.K., Sarma, J.D., 2020b. Histopathological imaging database for oral cancer analysis. Data Brief 29, 105114.
Sarrión Pérez, M.G., Bagán, J.V., Jiménez, Y., Margaix, M., Marzal, C., 2015. Utility of imaging techniques in the diagnosis of oral cancer. J. Cranio-Maxillo-Fac. Surg.: Off. Publ. Eur. Assoc. Cranio-Maxillo-Fac. Surg. 43 (9), 1880–1894.
Shah, J.P., Gil, Z., 2009. Current concepts in management of oral cancer–surgery. Oral Oncol. 45 (4–5), 394–401.
Shan, J., Jiang, R., Chen, X., Zhong, Y., Zhang, W., Xie, L., Cheng, J., Jiang, H., 2020. Machine learning predicts lymph node metastasis in early-stage oral tongue squamous cell carcinoma. J. Oral Maxillofac. Surg.: Off. J. Am. Assoc. Oral Maxillofac. Surg. 78 (12), 2208–2218.
Song, X., Yang, X., Narayanan, R., Shankar, V., Ethiraj, S., Wang, X., Duan, N., Ni, Y.H., Hu, Q., Zare, R.N., 2020. Oral squamous cell carcinoma diagnosed from saliva metabolic profiling. Proc. Natl. Acad. Sci. USA 117 (28), 16167–16173.
Stone, J.C., Glass, K., Munn, Z., Tugwell, P., Doi, S.A.R., 2020. Comparison of bias adjustment methods in meta-analysis suggests that quality effects modeling may have less limitations than other approaches. J. Clin. Epidemiol. 117, 36–45.
Stone, J.C., Gurunathan, U., Aromataris, E., Glass, K., Tugwell, P., Munn, Z., Doi, S.A.R., 2021. Bias assessment in outcomes research: the role of relative versus absolute approaches. Value Health: J. Int. Soc. Pharm. Outcomes Res. 24 (8), 1145–1149.
Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F., 2021. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer J. Clin. 71 (3), 209–249.
Sunny, S., Baby, A., James, B.L., Balaji, D., N, V.A., Rana, M.H., Gurpur, P., Skandarajah, A., D'Ambrosio, M., Ramanjinappa, R.D., Mohan, S.P., Raghavan, N., Kandasarma, U., N, S., Raghavan, S., Hedne, N., Koch, F., Fletcher, D.A., Selvam, S., Kollegal, M., N, P.B., Ladic, L., Suresh, A., Pandya, H.J., Kuriakose, M.A., 2019. A smart tele-cytology point-of-care platform for oral cancer screening. PLoS One 14 (11), e0224885.
Trajanovski, S., Shan, C., Weijtmans, P.J.C., de Koning, S.G.B., Ruers, T.J.M., 2021. Tongue tumor detection in hyperspectral images using deep learning semantic segmentation. IEEE Trans. Biomed. Eng. 68 (4), 1330–1340.
Vickers, A.J., Cronin, A.M., Elkin, E.B., Gonen, M., 2008. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC Med. Inform. Decis. Mak. 8, 53.
Wang, X., Yang, J., Wei, C., Zhou, G., Wu, L., Gao, Q., He, X., Shi, J., Mei, Y., Liu, Y., Shi, X., Wu, F., Luo, J., Guo, Y., Zhou, Q., Yin, J., Hu, T., Lin, M., Liang, Z., Zhou, H., 2020. A personalized computational model predicts cancer risk level of oral potentially malignant disorders and its web application for promotion of non-invasive screening. J. Oral Pathol. Med. 49 (5), 417–426.
Warin, K., Limprasert, W., Suebnukarn, S., Jinaporntham, S., Jantana, P., 2021. Automatic classification and detection of oral cancer in photographic images using deep learning algorithms. J. Oral Pathol. Med. 50 (9), 911–918.
Welikala, R.A., Remagnino, P., Lim, J.H., Chan, C.S., Rajendran, S., Kallarakkal, T.G., Zain, R.B., Jayasinghe, R.D., Rimal, J., Kerr, A.R., Amtha, R., Patil, K., Tilakaratne, W.M., Gibson, J., Cheong, S.C., Barman, S.A., 2020. Automated detection and classification of oral lesions using deep learning for early detection of oral cancer. IEEE Access 8, 132677–132693.
Whiting, P.F., Rutjes, A.W., Westwood, M.E., Mallett, S., Deeks, J.J., Reitsma, J.B., Leeflang, M.M., Sterne, J.A., Bossuyt, P.M., 2011. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann. Intern. Med. 155 (8), 529–536.
Yakob, M., Fuentes, L., Wang, M.B., Abemayor, E., Wong, D.T., 2014. Salivary biomarkers for detection of oral squamous cell carcinoma – current state and recent advances. Curr. Oral Health Rep. 1 (2), 133–141.
Yu, M., Yan, H., Xia, J., Zhu, L., Zhang, T., Zhu, Z., Lou, X., Sun, G., Dong, M., 2019. Deep convolutional neural networks for tongue squamous cell carcinoma classification using Raman spectroscopy. Photodiagn. Photodyn. Ther. 26, 430–435.
Zhao, Z., Li, Y., Wu, Y., Chen, R., 2020. Deep learning-based model for predicting progression in patients with head and neck squamous cell carcinoma. Cancer Biomark.: Sect. A Dis. Mark. 27 (1), 19–28.
Zhou, X., Hao, Y., Peng, X., Li, B., Han, Q., Ren, B., Li, M., Li, L., Li, Y., Cheng, G., Li, J., Ma, Y., Zhou, X., Cheng, L., 2021. The clinical potential of oral microbiota as a screening tool for oral squamous cell carcinomas. Front. Cell. Infect. Microbiol. 11.

Ibrahim Elmakaty is a 5th year medical student at Qatar University.

Mohamed Elmarasi is a 5th year medical student at Qatar University.

Ahmed Amarah is a 5th year medical student at Qatar University.

Ruba Abdo is a 5th year medical student at Qatar University.

Mohammed Imad Malki is the lead principle investigator and an Assistant Professor of Pathology in the College of Medicine at Qatar University.