



Title	Artificial intelligence for diagnosis and triage in oral cancer: a clinician-centered narrative review
Author(s)	Hiraoka, Shin-ichiro; Kawamura, Kohei; Akiyama, Ryo et al.
Citation	International Journal of Clinical Oncology. 2026, 31, p. 794-803
Version Type	VoR
URL	https://hdl.handle.net/11094/104579
rights	This article is licensed under a Creative Commons Attribution 4.0 International License.
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka



Artificial intelligence for diagnosis and triage in oral cancer: a clinician-centered narrative review

Shin-ichiro Hiraoka¹ · Kohei Kawamura¹ · Ryo Akiyama¹ · Yutaka Itakura¹ · Susumu Tanaka¹ · Narikazu Uzawa²

Received: 22 November 2025 / Accepted: 23 February 2026
© The Author(s) 2026

Abstract

Background Early diagnosis of oral squamous cell carcinoma (OSCC) remains challenging, with survival largely stage-dependent at presentation. Artificial intelligence (AI) promises to enhance detection and clinical decision-making across clinical photographs, radiology, optical imaging, and digital pathology.

Methods This narrative review synthesizes peer-reviewed PubMed-indexed English-language studies up to October 2025, prioritizing prospective designs, external validation, and clinically interpretable models. We focus on tasks relevant to clinicians: lesion triage from clinical images, prediction of nodal metastasis on CT/MRI/PET, margin assessment with optical modalities, and histopathology-based diagnosis/grading. We also discuss implementation issues: dataset shift, bias, and reporting standards.

Results In clinical photographs, deep learning achieves high diagnostic accuracy for OSCC and oral potentially malignant disorders (OPMD) classification in single-center studies and shows promising generalization with multi-site external testing, yet performance still degrades on out-of-distribution images and under real-world artifacts. In radiology, radiomics and deep learning models improve risk stratification and prediction of cervical nodal metastasis beyond conventional imaging, particularly with multimodal feature fusion. Optical methods such as hyperspectral spatial frequency domain imaging and OCT combined with AI show feasibility for intraoperative margin assessment and in-clinic triage. Digital pathology models on whole-slide images approach expert-level classification for OSCC diagnosis and are beginning to predict malignant transformation risk in oral epithelial dysplasia; however, rigorous prospective validation remains scarce. **Conclusion:** AI systems for OSCC are maturing and clinically oriented. Before routine adoption, studies must demonstrate external validity, clinician-in-the-loop performance, calibration, and impact on time-to-diagnosis and patient outcomes. Pragmatic trials and transparent reporting are essential to move beyond proof-of-concept into equitable clinical benefit.

✉ Shin-ichiro Hiraoka
hiraoka.shin-ichiro.dent@osaka-u.ac.jp

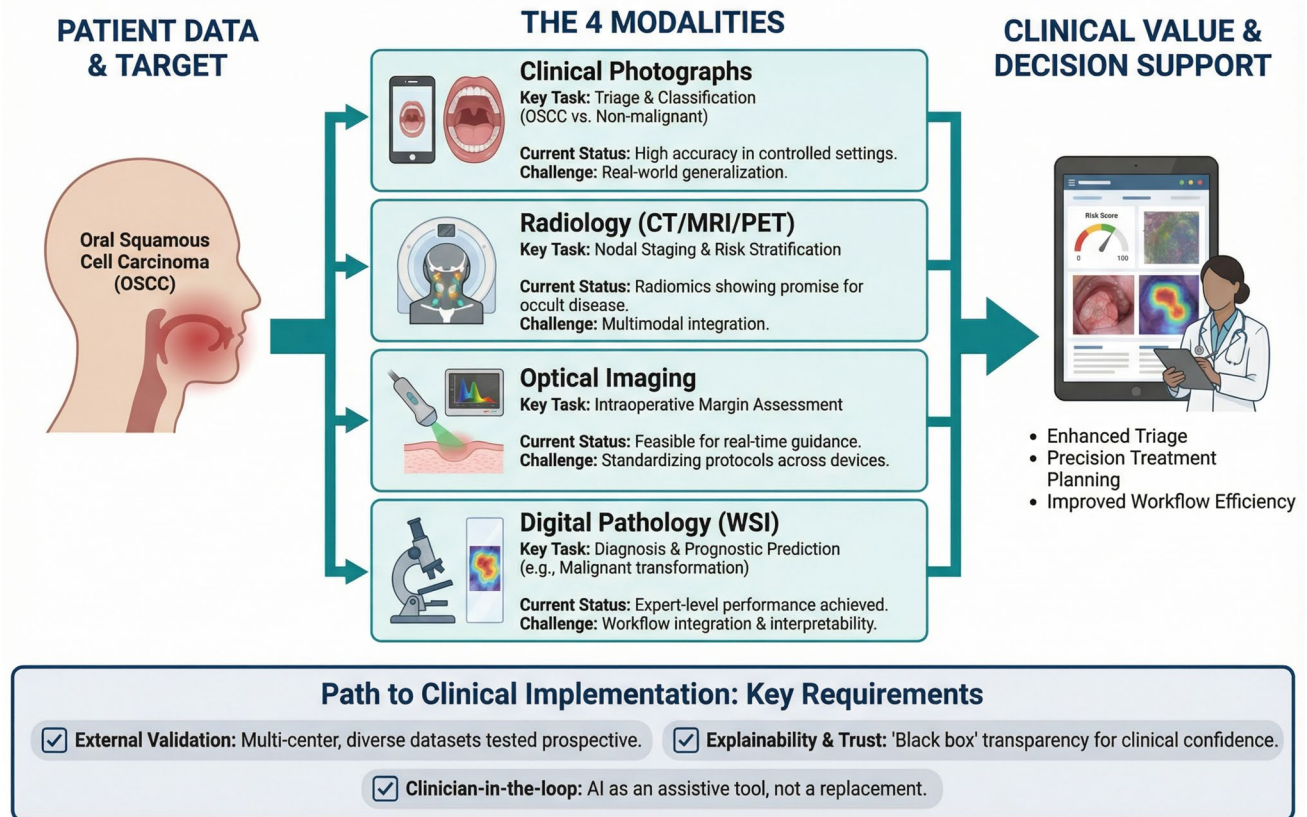
¹ Department of Oral and Maxillofacial Surgery, Graduate School of Dentistry, The University of Osaka, Osaka, Japan

² Department of Oral & Maxillofacial Oncology and Surgery, Graduate School of Dentistry, The University of Osaka, Osaka, Japan

Graphical Abstract

Artificial Intelligence for Diagnosis and Triage in Oral Cancer

A Clinician-Centred Narrative Review



Overview of AI-enabled diagnosis and triage in oral cancer across four modalities with validation priorities.

Keywords Oral squamous cell carcinoma · Artificial intelligence · Deep learning · Radiomics · Digital pathology · Lymph node metastasis

Introduction

Oral squamous cell carcinoma (OSCC) accounts for the vast majority of oral malignancies and remains a major global health burden. Outcomes are tightly linked to stage at diagnosis; delays at the level of patient presentation and health-system processes contribute to advanced-stage detection in many settings. AI techniques—particularly deep learning—have transformed pattern recognition in imaging and are rapidly permeating clinical oncology. Foundational work in general AI-enabled medicine and image classification laid much of the groundwork for domain-specific applications in oral oncology [1–3]. For oral cancer care, the central question is no longer whether AI can classify images in a laboratory, but whether it can shorten the diagnostic pathway,

improve risk stratification, and support decisions without amplifying inequities. Epidemiologic and diagnostic-delay contexts frame these challenges [4–8]. External validation is critical when translating artificial intelligence models into clinical practice. In this review, we define external validation as evaluation on data that are independent of model development and obtained from different institutions, time periods, acquisition settings, or patient populations. This matters because model performance often degrades under domain shift, and external validation is needed to support generalizability and transportability in real-world settings.

Scope and approach

This is a narrative (not systematic) review of PubMed-indexed, peer-reviewed, English-language studies published through October 2025. We searched PubMed using combinations of keywords related to oral squamous cell carcinoma, oral potentially malignant disorders, artificial intelligence, deep learning, radiomics, computed tomography, magnetic resonance imaging, positron emission tomography, optical coherence tomography, optical imaging, and digital pathology.

We focused on four clinician-facing tasks because they map to common decision points in care and have a growing body of peer-reviewed evidence: (i) triage from clinical photographs, (ii) prediction of cervical nodal metastasis on CT/MRI/PET, (iii) margin assessment and chair-side triage with optical modalities, and (iv) histopathology-based diagnosis or grading using whole-slide images.

We included original studies with clearly defined reference standards and quantitative evaluation. We prioritized studies that reported external validation or multi-site cohorts and those that provided clinically interpretable performance reporting, including discrimination, calibration, and error analysis. We excluded preprints and non-peer-reviewed sources to minimize citation drift. Because this is a narrative review, we did not aim for exhaustive coverage; instead, we selected representative studies to illustrate evidence maturity and implementation constraints across modalities.

Clinical photograph-based triage

A sustained line of work has explored AI classification from intraoral photographs captured by consumer-grade cameras. A landmark retrospective multicenter study showed that a convolutional neural network could identify oral cavity squamous cell carcinoma from clinical images with high accuracy and potential for telemedicine triage [9]. Subsequent reviews characterize common failure modes—lighting, blur, mucosal glare, and under-representation of darker mucosa—each of which degrades point-of-care performance [7, 10, 11]. Recent single- and multi-center datasets, including transformer-based approaches, reported improved generalization, and external testing for OSCC and oral potentially malignant disorders (OPMD) differentiation [9, 12, 13]. Complementing these image-centered models, a reader-style evaluation using an LLM (GPT-4) on clinical photographs—with and without clinical history—showed better performance when history was available, underscoring clinician-in-the-loop workflows [14]. In parallel, smaller pilot studies using object-detection frameworks demonstrate feasibility but need larger cohorts and external validation [15]. Methods tailored to early disease presentation continue to appear and may help address subtle lesions [16].

Meta-analytic evidence suggests that AI-assisted image assessment achieves high pooled accuracy across imaging tools, though heterogeneity and reference-standard differences remain [17, 18]. Representative clinical photograph studies and their validation characteristics are summarized in Table 1.

Cross-sectional radiology: nodal staging and risk stratification

In radiology, the main clinical aim is preoperative staging and risk stratification that informs surgical planning, neck management, and adjuvant therapy. Most oral cancer AI work, therefore, targets prediction of cervical nodal metastasis and related adverse features, and it secondarily supports prognostication and more standardized image interpretation. Figure 1 summarizes a typical radiology AI workflow, from image acquisition and feature extraction to model evaluation and validation.

Radiomics and deep learning applied to CT/MRI have repeatedly shown incremental value for predicting nodal metastasis, extracapsular spread, and occult nodal disease beyond radiologist assessment alone [19–21]. Models that integrate tumor textural features with nodal morphology and clinical variables tend to calibrate better than image-only approaches. PET-based radiomics adds prognostic information for local control and survival, and node-level classifiers can outperform semi-quantitative features [22, 23]. PET radiomics has also been used to predict histologic grade preoperatively in OSCC [24]. Beyond nodal staging and histologic grade prediction, MRI radiomics has also been explored as a surrogate of the tumor immune microenvironment; in a single-center pilot on oral tongue SCC, a machine learning model estimated tumor-infiltrating lymphocyte levels from preoperative MRI, suggesting feasibility but requiring external validation [25]. A recent meta-analysis indicates that hybrid PET/MRI performs at least comparably with PET/CT for primary and nodal evaluation, with overlapping confidence intervals in head-to-head analyses [26]. For broader clinical positioning of PET/CT in HNSCC, contemporary narrative reviews remain useful [27]. However, most reports remain retrospective and heterogeneous in endpoints and validation strategies, which limits transportability across sites and scanners. Table 2 summarizes representative studies and highlights gaps in external validation and clinically meaningful endpoints.

Optical imaging for margin assessment and chair-side triage

Optical methods seek to bridge the gap between coarse white-light inspection and definitive histology. Hyperspectral spatial frequency domain imaging (SFDI) analyzed with

Table 1 Representative studies using clinical photographs for OSCC/OPMD classification and triage

First author (year); Journal	Country/centers	Design	N (images/patients; split)	Task	External validation	Primary performance (external test preferred)	Comparator
Fu (2020)[9]	China; 11 centers	Retrospective, multi-center	44,409 images; internal $n = 401$, external $n = 402$	Detect OSCC from clinical photographs	Yes	External: AUC 0.935 (95% CI 0.910–0.957), Sens 89.6%, Spec 80.6%; internal: AUC 0.983 (0.973–0.991), Sens 94.9%, Spec 88.7%	7 oral-cancer specialists (comparable)
Kouketsu (2024)[12]	Japan; single center (outpatient clinic)	Retrospective image-based	1,043 images from 424 patients	Detect OSCC (and OSCC + leukoplakia) from clinical photos	No (not reported)	OSCC detection: Sens 93.9%, Spec 81.2%; OSCC + leukoplakia: Sens 83.7%, Spec 81.2%	—
Keser (2024); [15]	Türkiye; single center	Retrospective pilot (YOLOv5)	65 images; split 53/6/6 (train/val/test)	Detect oral cancer in clinical photos	No	Test: F1 0.667, Sens 0.667, precision 0.667	—
Schmidl (2025); [14]	Germany; single center	Prospective reader study (LLM assistant)	45 patients (histology proven)	AI assistant (GPT-4o) classifies OSCC vs non-OSCC from photos $\pm$ history	No	With history: OSCC Sens 100%, Spec 88.2%; leukoplakia Sens 93.3%, Spec 96.7%	Human readers (context-aware)

Reported metrics are as published in the source articles

AUC area under the ROC curve, *Sens* sensitivity, *Spec* specificity, *95% CI* 95% confidence interval, *OSCC* oral cavity squamous cell carcinoma, *OSCC* oral squamous cell carcinoma, *OED* oral epithelial dysplasia, *OED MT* malignant transformation, *EV* external validation, *LLM* large language model

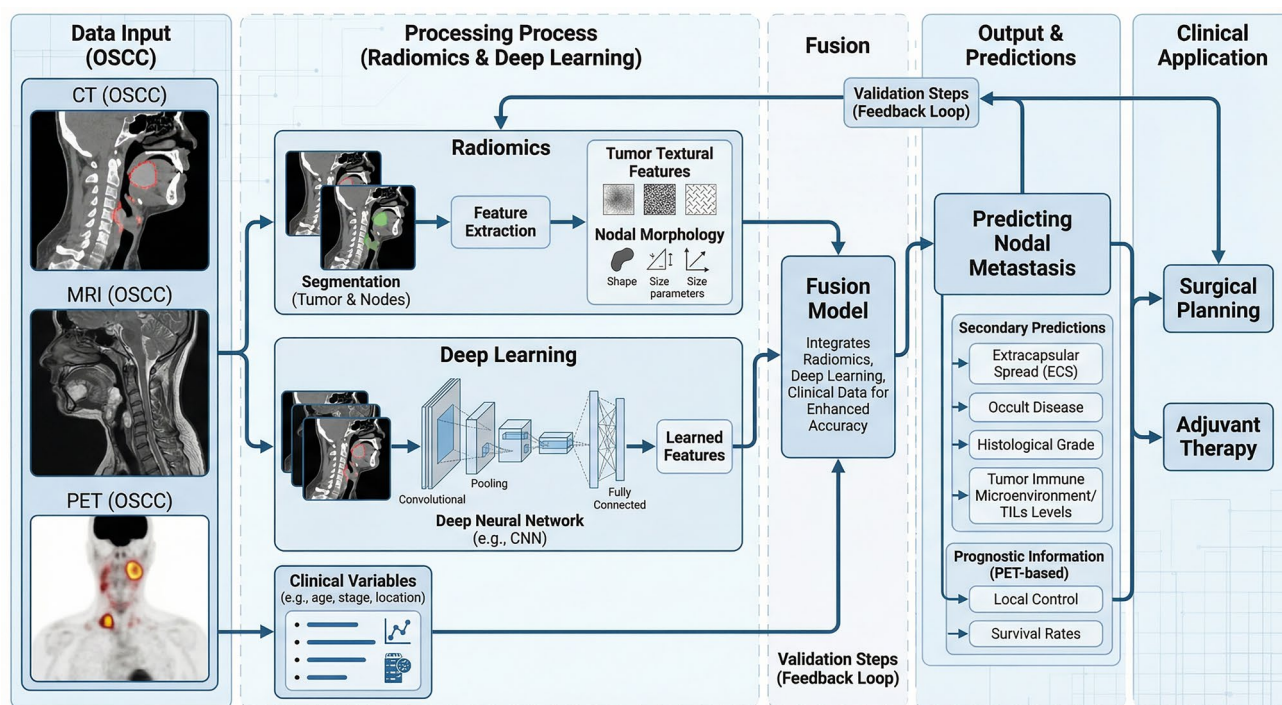


Fig. 1 Radiology pipeline

deep learning has been explored in preclinical and ex vivo settings to estimate subsurface margin thickness, suggesting feasibility for intraoperative margin assessment in the future [28]. Optical coherence tomography (OCT) meta-evidence and systematic reviews support high diagnostic accuracy for oral lesions, and AI-assisted OCT shows promise for chair-side triage when histology is unavailable [29]. These modalities are attractive precisely because they do not require large laboratories or complex post-processing at the bedside—yet their impact depends on probe ergonomics in the oral cavity, training curves, and how confidently clinicians can act on model outputs while avoiding unnecessary biopsies.

Digital pathology: from diagnosis to prediction

Whole-slide image (WSI) analysis is the most mature AI application in OSCC. Figure 2 illustrates the end-to-end digital pathology workflow for OSCC diagnosis/grading and OED malignant transformation risk modeling. Multiple groups have reported approaching expert-level performance and augmenting pathologists in the histopathologic diagnosis of OSCC versus benign epithelium, and have begun to quantify tumor budding, grade, and other histologic predictors [30–32]. A study showed that deep learning classifiers can augment pathologists' performance, particularly for challenging borderline regions [31]. Importantly, AI is starting to address prediction tasks that lock tightly to clinical decisions. Recent work in oral epithelial dysplasia (OED)

proposes AI pipelines to predict malignant transformation from baseline WSIs—conceptually akin to risk calculators in other cancers. A 2025 study reported external validation for an OED transformation model, a critical milestone for moving beyond retrospective case–control classification [33]. Updates focused specifically on OED in digital pathology complement these advances [34]. Representative whole-slide image studies in OSCC and OED are summarized in Table 3

Integrative perspectives and clinician-in-the-loop design

This section synthesizes evidence maturity and implementation considerations across modalities. To align the narrative with the evidence maturity map, we treat three study-design dimensions as the core indicators of clinical credibility: external validation, prospective evaluation, and clinically meaningful endpoints. In addition, we discuss cross-cutting factors that influence interpretability and reproducibility, including multimodal modeling, dataset diversity and equity, sample size, multi-center recruitment, transparency of pre-processing and model development, and appropriateness of evaluation metrics. Three patterns recur. First, external validation remains the exception; when present, performance typically drops relative to internal testing, as shown across clinical photographs, digital pathology, and radiology [9, 19, 33]. Second, multimodal approaches often improve

Table 2 Representative radiology studies applying radiomics/deep learning for nodal staging, prognostication, and PET-based risk assessment

First author (year); Journal	Modality/target	Design	Cohort (split/EV)	Primary metric (external test preferred)	Clinical endpoint
Lan (2024); [19]	MRI radiomics + DL; predict occult cervical LNM in early OSCC/OPSCC	Retrospective; multi-center with two external validation sets	$n = 319$; EV set1/2 AUCs reported	AUC train 0.928; test 0.878; external1 0.796; external2 0.834 (ResNet50 fusion)	Pre-op OCLNM prediction and prognosis
Bianconi (2024); [23]	FDG-PET radiomics; nodal status in HNC	Retrospective; per-node analysis	27 patients; 51 LNs; LOOCV	LR: PETBase AUC 0.840 vs PETRad AUC 0.880; Sens 68.0% $\rightarrow$ 72.0%, Spec 89.5% $\rightarrow$ 90.0%	Pathologic LN classification
Al-Ibraheem (2024) [26]; (meta-analysis)	FDG-PET/MRI (hybrid) for HNSCC	Systematic review and meta-analysis (12 studies)	638 patients overall	Pooled sensitivity/specificity: primary disease 93%/95% (per-patient); locoregional 93%/96%; nodal disease 89%/98% (per-lesion)	Diagnostic accuracy vs PET/CT or MRI
Martens (2020); [22]	FDG-PET radiomics; outcomes in LA-HNSCC	Retrospective; train/validation	Train $n = 103$; Val $n = 71$	Concordance index (validation): 0.645 (recurrence), 0.627 (metastasis)	Recurrence/distant metastasis/OS prediction
Nikkuni (2024); [24]	FDG-PET radiomics; histological grade of OSCC	Retrospective; supervised ML	Single center; retrospective $n = 191$	AUCs: LR 0.72, SVM 0.71, RF 0.84, NB 0.74, KNN 0.73	Grade prediction (pre-op)

LNM lymph node metastasis, OCLNM occult cervical lymph node metastasis, HNSCC head and neck squamous cell carcinoma, AUC area under ROC, EV external validation, LR logistic regression, RF random forest, SVM support vector machine

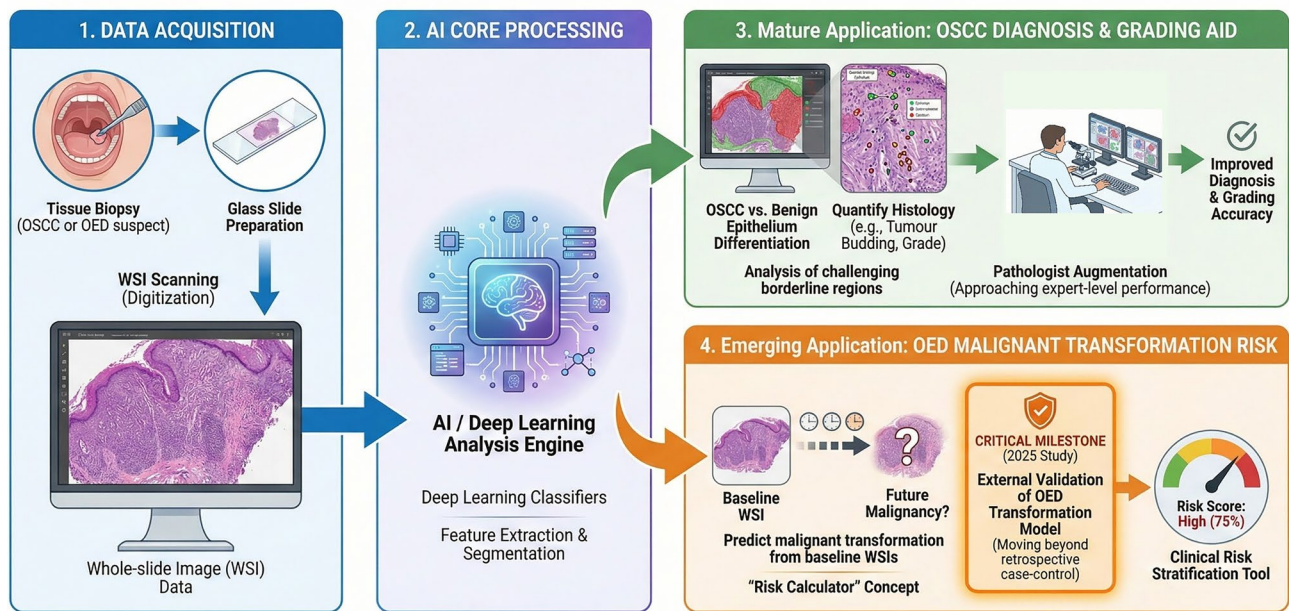


Fig. 2 Digital pathology pipeline. End-to-end digital pathology workflow for OSCC diagnosis/grading and OED malignant transformation risk modeling

discrimination and may yield more stable calibration than single-source models—mirroring how clinicians synthesize data across domains [19, 22]. Third, equity and generalizability must be designed from the start: datasets should reflect the full spectrum of oral mucosal pigmentation, dental restorations and prostheses, and the diversity of lesion mimics seen in general practice, not solely tertiary centers [7, 10]. Transparent reporting aligned with the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) [37, 38] and STARD (Standards for Reporting Diagnostic Accuracy Studies) [39, 40] guidelines, including their AI-oriented extensions, and prospective impact evaluations anchored to outcomes, such as time-to-biopsy, stage shift, and avoidable re-excisions, are essential.

These patterns are summarized qualitatively in Table 4, which provides a directional snapshot of evidence maturity by modality and study design dimension. In brief, clinical photographs and digital pathology show emerging strength for external validation, whereas prospective designs and clinical impact endpoints remain scarce across all modalities; radiology (CT/MRI/PET) shows a similar profile, and optical imaging currently has sparse evidence on all three axes [17, 18, 29, 34].

From an implementation standpoint, two challenges dominate: domain shift across devices, acquisition protocols, and clinical settings, and the need for trustworthy model behavior at the point of care. Domain shift arises from differences in camera type and lighting in clinical photographs,

scanner and reconstruction settings in radiology, acquisition constraints in optical imaging, and staining and scanning pipelines in digital pathology. Mitigation requires standardized acquisition where feasible, robust augmentation, and site-aware external validation.

Explainability methods may support clinician acceptance, but post hoc attributions do not guarantee faithful explanations. Therefore, deployment should emphasize calibrated uncertainty, explicit failure-mode reporting, and monitoring for performance drift after rollout [41].

Practical guidance for clinicians and services

For clinics exploring adoption, we recommend a staged approach. Begin with adjunctive triage applications for which the action is uncontroversial (e.g., flagging high-risk lesions for urgent assessment, standardizing photographic capture). Use shadow-mode audits to measure real-world false positive/negative profiles before allowing the system to influence decisions. For radiology services, prioritize models with open documentation, proven external datasets, and locked algorithms; integrate them into reporting templates that expose model uncertainty and rationales. For pathology departments, invest in robust WSI pipelines and metadata governance first; model selection comes second. Finally, define exit criteria—the thresholds at which an AI tool will be rolled back if drift, bias, or calibration failure emerges. Regulatory status should be considered explicitly. AI tools intended for diagnostic use may require medical-device

Table 3 Digital pathology studies

First author (year); Journal	Specimen/task	Design	Cohort	External validation	Primary metric	Clinical endpoint
Shephard (2024); [35]	WSI (OED); malignant transformation prediction (OMTscore)	Retrospective; internal CV + 2 external cohorts	Internal $n = 193$; external $n = 89$ (2 sites)	Yes (2 centers)	External AUROC 0.75 (recall 0.92); C-index 0.60 ($p = 0.02$)	Prognostic stratification for OED (MT risk)
Shephard (2025); [33]	WSI (OED); malignant transformation prediction (ODYN)	Development + external validation (3 centers; 3 scanners)	Train 358 OED + 105 controls; external 108 OED	Yes (3 centers)	AUROC 0.73 (external); C-index 0.63 (external)	Prognostic risk scoring for OED
Sukegawa (2024); [36]	Oral exfoliative cytology; CNN classification	Retrospective; repeated CV, probabilistic labels	14,535 patch images; 6 WSIs (LBC)	No	AUC up to 0.991; accuracy up to 0.988 across cases; prior baseline accuracy 0.902 (patch-level)	Screening-adjunct classification

Representative whole-slide image studies in OSCC and OED, including diagnostic classification and risk prediction
MT risk malignant transformation risk

Table 4 Evidence maturity matrix (qualitative)

Modality	External validation	Prospective design	Clinical impact endpoints
Clinical photos	○	—	—
Radiology (CT/MRI/PET)	○	—	—
Optical imaging	—	—	—
Digital pathology	○	—	—

Qualitative maturity of the evidence by modality and design dimension

Design dimensions: external validation, prospective design, clinical impact endpoints. Legend: ○ emerging, — sparse. Ratings reflect peer-reviewed literature up to October 2025 and are intended as a directional summary [9, 14, 17–20, 23, 27, 29, 34]

regulatory authorization depending on jurisdiction. Within the studies included in this review, regulatory authorization and routine clinical deployment are not consistently described; this gap likely reflects limited multi-site validation, workflow integration constraints, and evidentiary requirements for authorization. We, therefore, distinguish research-grade prototypes from evidence that supports clinical deployment when summarizing each modality. Table 5 provides a concise checklist of minimum requirements and safeguards for clinical deployment.

Limitations of the evidence

Most studies remain retrospective and single-center; many lack pre-registered protocols, clinical impact measures, or detailed error analyses. Performance is often reported as point estimates without decision-curve analyses or cost-effectiveness. Publication bias and dataset reuse complicate pooled estimates. These limitations should temper enthusiasm and motivate stronger methodology rather than discourage exploration.

Conclusion

AI for oral cancer diagnosis has progressed from promising prototypes to clinically meaningful tools across images acquired by clinicians, radiologists, and pathologists. The evidence base now supports adjunctive use for triage and risk stratification, with emerging signals for predictive modeling in OED. The next decisive step is not a marginal AUC gain but trustworthy, externally validated systems embedded in clinical workflows that shorten diagnostic delays and improve patient outcomes.

Table 5 Minimum requirements and safeguards for clinical deployment of AI systems in oral oncology

Domain	Requirement/safeguard (minimum requirement)	Rationale/specifics
Phased rollout (staged approach)	Begin with adjunctive triage applications	e.g., Flagging high-risk lesions for urgent assessment or standardizing photographic capture
Pre-deployment validation	Use 'shadow-mode' audits before live deployment	Measure real-world false positive/negative profiles before the system is allowed to influence clinical decisions
Model vetting (radiology)	Prioritize models with open documentation, proven external datasets, and locked algorithms	Ensure transparency and validated performance
Model vetting (pathology)	Invest in robust WSI (whole-slide image) pipelines and metadata governance first	Model selection is secondary to the quality of the underlying digital infrastructure
Clinical integration	Integrate AI outputs into reporting templates that expose model uncertainty and rationales	Clinicians must be able to see the AI's confidence and (if possible) its reasoning
Ongoing governance and monitoring	Define clear 'exit criteria' for rolling back the tool	Must continuously monitor for performance drift, bias, or calibration failure after deployment

Acknowledgements The authors thank the Japan Society of Clinical Oncology (JSCO) for the invitation to write this review. We also thank colleagues who contributed discussions and feedback on AI applications in oral oncology.

Funding Open Access funding provided by The University of Osaka. This work was supported by an honorarium from the Japan Society of Clinical Oncology (JSCO).

Data availability Not applicable.

Declarations

Conflict of interest Hiraoka received an honorarium from the Japan Society of Clinical Oncology (JSCO) for writing this invited review. The remaining authors declare no conflicts of interest.

Ethical approval Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Esteva A, Kuprel B, Novoa RA et al (2017) Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542(7639):115–118. <https://doi.org/10.1038/nature21056>
2. Ilhan B, Lin K, Guneri P et al (2020) Improving oral cancer outcomes with imaging and artificial intelligence. *J Dent Res* 99(3):241–248. <https://doi.org/10.1177/0022034520902128>
3. Mahmood H, Shaban M, Rajpoot N et al (2021) Artificial intelligence-based methods in head and neck cancer diagnosis: an overview. *Br J Cancer* 124(12):1934–1940. <https://doi.org/10.1038/s41416-021-01386-x>
4. Sung H, Ferlay J, Siegel RL et al (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 71(3):209–249. <https://doi.org/10.3322/caac.21660>
5. González-Moles M, Aguilar-Ruiz M, Ramos-García P (2022) Challenges in the early diagnosis of oral cancer, evidence gaps and strategies for improvement: a scoping review of systematic reviews. *Cancers (Basel)*. <https://doi.org/10.3390/cancers14194967>
6. Johnson NW, Jayasekara P, Amarasinghe AA (2011) Squamous cell carcinoma and precursor lesions of the oral cavity: epidemiology and aetiology. *Periodontol* 2000 57(1):19–37. <https://doi.org/10.1111/j.1600-0757.2011.00401.x>
7. Vinay V, Jodalli P, Chavan MS et al (2025) Artificial intelligence in oral cancer: a comprehensive scoping review of diagnostic and prognostic applications. *Diagnostics (Basel)*. <https://doi.org/10.3390/diagnostics15030280>
8. Pires FR, Ramos AB, Oliveira JB et al (2013) Oral squamous cell carcinoma: clinicopathological features from 346 cases from a single oral pathology service during an 8-year period. *J Appl Oral Sci* 21(5):460–467. <https://doi.org/10.1590/1679-775720130317>
9. Fu Q, Chen Y, Li Z et al (2020) A deep learning algorithm for detection of oral cavity squamous cell carcinoma from photographic images: a retrospective study. *EclinicalMedicine* 27:100558. <https://doi.org/10.1016/j.eclinm.2020.100558>
10. Chu CS, Lee NP, Ho JWK et al (2021) Deep learning for clinical image analyses in oral squamous cell carcinoma: a review. *JAMA Otolaryngol Head Neck Surg* 147(10):893–900. <https://doi.org/10.1001/jamaoto.2021.2028>
11. Warin K, Suebnukarn S (2024) Deep learning in oral cancer- a systematic review. *BMC Oral Health* 24(1):212. <https://doi.org/10.1186/s12903-024-03993-5>
12. Kouketsu A, Doi C, Tanaka H et al (2024) Detection of oral cancer and oral potentially malignant disorders using artificial intelligence-based image analysis. *Head Neck* 46(9):2253–2260. <https://doi.org/10.1002/hed.27843>
13. Song B, Kc DR, Yang RY et al (2024) Classification of mobile-based oral cancer images using the vision transformer and the swin transformer. *Cancers (Basel)*. <https://doi.org/10.3390/cancers16050987>

14. Schmidl B, Hütten T, Pigorsch S et al (2025) Artificial intelligence for image recognition in diagnosing oral and oropharyngeal cancer and leukoplakia. *Sci Rep* 15(1):3625. <https://doi.org/10.1038/s41598-025-85920-4>
15. Keser G, Pekiner FN, Bayrakdar İ et al (2024) A deep learning approach to detection of oral cancer lesions from intra oral patient images: a preliminary retrospective study. *J Stomatol Oral Maxillofac Surg* 125(12 Suppl 2):101975. <https://doi.org/10.1016/j.jormas.2024.101975>
16. Liu P, Bagi K (2025) A tailored deep learning approach for early detection of oral cancer using a 19-layer CNN on clinical lip and tongue images. *Sci Rep* 15(1):23851. <https://doi.org/10.1038/s41598-025-07957-9>
17. Ren C, Wang C, Lin R et al (2025) The diagnostic value of artificial intelligence in oral squamous cell carcinoma: a systematic review and meta-analysis. *J Stomatol Oral Maxillofac Surg* 126(5s):102429. <https://doi.org/10.1016/j.jormas.2025.102429>
18. Li J, Kot WY, McGrath CP et al (2024) Diagnostic accuracy of artificial intelligence assisted clinical imaging in the detection of oral potentially malignant disorders and oral cancer: a systematic review and meta-analysis. *Int J Surg* 110(8):5034–5046. <https://doi.org/10.1097/js9.0000000000001469>
19. Lan T, Kuang S, Liang P et al (2024) MRI-based deep learning and radiomics for prediction of occult cervical lymph node metastasis and prognosis in early-stage oral and oropharyngeal squamous cell carcinoma: a diagnostic study. *Int J Surg* 110(8):4648–4659. <https://doi.org/10.1097/js9.0000000000001578>
20. Deng C, Hu J, Tang P et al (2024) Application of CT and MRI images based on artificial intelligence to predict lymph node metastases in patients with oral squamous cell carcinoma: a subgroup meta-analysis. *Front Oncol* 14:1395159. <https://doi.org/10.3389/fonc.2024.1395159>
21. Illimoottil M, Ginat D (2023) Recent advances in deep learning and medical imaging for head and neck cancer treatment: MRI, CT, and PET scans. *Cancers (Basel)*. <https://doi.org/10.3390/cancers15133267>
22. Martens RM, Koopman T, Noij DP et al (2020) Predictive value of quantitative (18)F-FDG-PET radiomics analysis in patients with head and neck squamous cell carcinoma. *EJNMMI Res* 10(1):102. <https://doi.org/10.1186/s13550-020-00686-2>
23. Bianconi F, Salis R, Fravolini ML et al (2024) Radiomics features from positron emission tomography with [(18)F] fluorodeoxyglucose can help predict cervical nodal status in patients with head and neck cancer. *Cancers (Basel)*. <https://doi.org/10.3390/cancers16223759>
24. Nikkuni Y, Nishiyama H, Hayashi T (2024) Prediction of histological grade of oral squamous cell carcinoma using machine learning models applied to (18)F-FDG-PET radiomics. *Biomedicines*. <https://doi.org/10.3390/biomedicines12071411>
25. Ren J, Yang G, Song Y et al (2024) Machine learning-based MRI radiomics for assessing the level of tumor infiltrating lymphocytes in oral tongue squamous cell carcinoma: a pilot study. *BMC Med Imaging* 24(1):33. <https://doi.org/10.1186/s12880-024-01210-x>
26. Al-Ibraheem A, Abdulkadir A, Herrmann K et al (2024) Diagnostic accuracy of [(18)F]FDG PET/MRI in head and neck squamous cell carcinoma: a systematic review and metaanalysis. *J Nucl Med* 65(10):1533–1539. <https://doi.org/10.2967/jnumed.124.268049>
27. Caldarella C, De Risi M, Massacesi M et al (2024) Role of (18) F-FDG PET/CT in head and neck squamous cell carcinoma: current evidence and innovative applications. *Cancers (Basel)*. <https://doi.org/10.3390/cancers16101905>
28. Kurosawa H, Won NJ, Wunder JB et al (2025) Deep learning-enabled fluorescence imaging for oral cancer margin classification in preclinical models. *J Biomed Opt* 30(Suppl 3):S34109. <https://doi.org/10.1117/1.Jbo.30.S3.S34109>
29. Jerjes W, Stevenson H, Ramsay D et al (2024) Enhancing oral cancer detection: a systematic review of the diagnostic accuracy and future integration of optical coherence tomography with artificial intelligence. *J Clin Med*. <https://doi.org/10.3390/jcm13195822>
30. Yang SY, Li SH, Liu JL et al (2022) Histopathology-based diagnosis of oral squamous cell carcinoma using deep learning. *J Dent Res* 101(11):1321–1327. <https://doi.org/10.1177/00220345221089858>
31. Sukegawa S, Ono S, Tanaka F et al (2023) Effectiveness of deep learning classifiers in histopathological diagnosis of oral squamous cell carcinoma by pathologists. *Sci Rep* 13(1):11676. <https://doi.org/10.1038/s41598-023-38343-y>
32. Deif MA, Attar H, Amer A et al (2022) Diagnosis of oral squamous cell carcinoma using deep neural networks and binary particle swarm optimization on histopathological images: an AIoMT approach. *Comput Intell Neurosci* 2022:6364102. <https://doi.org/10.1155/2022/6364102>
33. Shephard AJ, Mahmood H, Raza SEA et al (2025) Development and validation of an artificial intelligence-based pipeline for predicting oral epithelial dysplasia malignant transformation. *Commun Med* 5(1):186. <https://doi.org/10.1038/s43856-025-00873-z>
34. Alajaji SA, Khoury ZH, Jessri M et al (2024) An update on the use of Artificial Intelligence in digital pathology for oral epithelial dysplasia research. *Head Neck Pathol* 18(1):38. <https://doi.org/10.1007/s12105-024-01643-4>
35. Shephard AJ, Bashir RMS, Mahmood H et al (2024) A fully automated and explainable algorithm for predicting malignant transformation in oral epithelial dysplasia. *NPJ Precis Oncol* 8(1):137. <https://doi.org/10.1038/s41698-024-00624-8>
36. Sukegawa S, Tanaka F, Nakano K et al (2024) Training high-performance deep learning classifier for diagnosis in oral cytology using diverse annotations. *Sci Rep* 14(1):17591. <https://doi.org/10.1038/s41598-024-67879-w>
37. Collins GS, Moons KGM, Dhiman P et al (2024) TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 385:e078378. <https://doi.org/10.1136/bmj-2023-078378>
38. Moons KG, Altman DG, Reitsma JB et al (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 162(1):W1–73. <https://doi.org/10.7326/m14-0698>
39. Sounderajah V, Guni A, Liu X et al (2025) The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. *Nat Med* 31(10):3283–3289. <https://doi.org/10.1038/s41591-025-03953-8>
40. Bossuyt PM, Reitsma JB, Bruns DE et al (2015) STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 351:h5527. <https://doi.org/10.1136/bmj.h5527>
41. Pushparathi VPG, Narayan SRS, R SP et al (2025) Histopathological image analysis and enhanced diagnostic accuracy explainability for oral cancer detection. *Pol J Pathol* 76(2):120–130. <https://doi.org/10.5114/pjp.2025.153973>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.