Archives of Medical Reports

3(4) (2026) 8-16

OPEN ACCESS

REVIEW ARTICLE

Artificial Intelligence and Digital Epidemiology for Population-Based Oral Cancer Risk Prediction

Wisdom Azumah [1,*], Jananee Bharkavi [2,*], Jocelyn Wensel [3,*], Laxmi Pradhan [4,*]

[1] Department of Biochemistry and Biotechnology, Kwame Nkrumah University of Science and Technology (KNUST), Kumasi, Ashanti Region, Ghana.

[2] University of Glasgow, Glasgow G12 8QQ, United Kingdom.

[3] International University of Health Sciences, Basseterre, Saint Kitts and Nevis.

[4] Institute of Medicine, Tribhuvan University, Kathmandu, Nepal.

***Corresponding Author:** Wisdom Azumah, Jananee Bharkavi, Jocelyn Wensel, Laxmi Pradhan
[1]Department of Biochemistry and Biotechnology, Kwame Nkrumah University of Science and Technology (KNUST), Kumasi, Ashanti Region, Ghana
[2] University of Glasgow, Glasgow G12 8QQ, United Kingdom
[3] International University of Health Sciences, Basseterre, Saint Kitts and Nevis
[4] Institute of Medicine, Tribhuvan University, Kathmandu, Nepal
Email: azumahwisdom854@gmail.com, jananeebharkavi@gmail.com, jocelynw@iuhs.edu, laxmipradhan1229@gmail.com

Article info
Received: 19 May 2026
Accepted: 19 July 2026

Keywords: Artificial Intelligence; Digital Epidemiology; Oral Cancer; Digital Health.

How to cite this article: Wisdom Azumah, Jananee Bharkavi, Jocelyn Wensel, Laxmi Pradhan (2026). Artificial Intelligence and Digital Epidemiology for Population-Based Oral Cancer Risk Prediction, 3(4), 8-16. https://doi.org/10.26524/amr.2026.4.2

Abstract

Oral cancer remains a major public health problem because many patients are diagnosed after substantial disease progression. Oral squamous cell carcinoma represents the major histological form of oral cancer. Conventional oral examination remains essential for early recognition. However, screening programmes are difficult to implement uniformly across populations because specialist resources are unevenly distributed. Digital epidemiology provides an opportunity to collect population-level information through electronic health records. Artificial intelligence can analyse these heterogeneous datasets and identify patterns associated with disease risk. Recent studies have demonstrated the potential of machine learning for oral cancer risk prediction. Deep learning has also produced high diagnostic performance using clinical photographs and histopathological images. Smartphone-based systems are beginning to extend AI-assisted screening into community settings. However, current evidence remains dominated by retrospective studies. Many datasets are derived from single institutions and contain limited demographic diversity. External validation remains insufficient. Differences in image acquisition, disease prevalence, and population characteristics can influence model performance. This review discusses how artificial intelligence and digital epidemiology can be integrated into a population-based oral cancer prediction framework.

1. Introduction

Oral cancer is an important global malignancy with considerable effects on mortality, morbidity, and quality of life. The disease includes malignancies arising from the lip and oral cavity. Global cancer estimates have demonstrated a substantial burden of oral cancer with the disease particularly affecting regions where tobacco, areca nut, and betel quid exposure are common (Rumgay et al., 2024a; Sung et al., 2021) The distribution of oral cancer is therefore not uniform. Differences in behavioral exposures and preventive practices produce substantial variation between populations. These differences make oral cancer an important target for population-based risk assessment. The clinical outcome of oral cancer is strongly associated with the stage at diagnosis. Early lesions may be amenable to treatment with a better chance of preserving oral function and reducing treatment-related morbidity. Advanced disease can require extensive surgery, radiotherapy, chemotherapy, or combinations of

© The Author(s). 2026 Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and non-commercial reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

these approaches. Late diagnosis remains a major concern in many populations. Conventional visual examination remains the foundation of oral cancer screening. Oral potentially malignant disorders (OPMD) provide an important opportunity for early intervention. OPMDs represent a heterogeneous group of oral conditions associated with an increased risk of malignant transformation (Warnakulasuriya et al., 2021). Leukoplakia, erythroplakia, oral submucous fibrosis, and other potentially malignant conditions do not have identical biological behavior. Some lesions remain stable for prolonged periods. Other lesions develop dysplasia and eventually progress to invasive cancer. Histopathological examination remains central to risk assessment, but dysplasia grading can show interobserver variation (Gupta et al., 2020) Clinical appearance alone also cannot reliably predict the future behavior of every lesion. These characteristics create a need for models that can integrate several clinical and pathological variables rather than depending on a single observation.

Artificial intelligence has emerged as an important tool for this purpose. Machine learning algorithms can identify associations between demographic variables, clinical characteristics, and disease outcomes. Deep learning can identify complex features in clinical photographs and histopathological images. Natural language processing can potentially extract risk information from electronic clinical records. These approaches provide an opportunity to move from conventional single-variable risk assessment toward multidimensional prediction. The concept is particularly relevant to oral cancer because disease risk is influenced by multiple interacting factors rather than a single exposure. The current oral cancer AI literature has mainly focused on diagnosis. The recent study has reviewed 54 deep learning studies and found that 51 focused on diagnosis while only three addressed prognostic prediction (Warin & Suebnukarn, 2024). Diagnostic studies used histopathological images, clinical photographs, computed tomography, and other medical imaging modalities. Reported performance was generally high, although substantial heterogeneity existed between studies. A more recent systematic review and meta-analysis published in 2026 included 16 AI studies and reported pooled sensitivity of 0.90, specificity of 0.89, and accuracy of 0.89 (Soleimani Sardou et al., 2026). Image-based models showed pooled sensitivity of 0.94 and specificity of 0.93. However, heterogeneity was extremely high, with an I^2 value of 100%, indicating that the pooled estimates should not be interpreted as universally applicable clinical performance values. A population-based AI system could begin with demographic and behavioral risk factors. Individuals identified as high risk could receive targeted oral examination. Smartphone imaging could then assist with lesion assessment. Suspicious cases could be referred for specialist examination and biopsy. Histopathology could provide definitive diagnosis. Following diagnosis, AI could assist with prognostic prediction and treatment-related risk assessment. Such an approach would transform AI from an isolated diagnostic tool into a component of an integrated oral cancer prevention pathway. The objective of this review is therefore to examine the emerging relationship between artificial intelligence and digital epidemiology in population-based oral cancer risk prediction.

2. Population Risk Landscape and the Need for Digital Epidemiology

Oral cancer risk is strongly influenced by preventable behavioral exposures. Tobacco remains one of the most important risk factors. Smoking exposes the oral mucosa to numerous carcinogenic compounds. Smokeless tobacco can produce prolonged local exposure and is particularly important in populations where chewing tobacco products are commonly used. Areca nut represents another major risk factor and it could be consumed alone or within betel quid preparations. Tobacco may also be added to these products. The importance of these exposures is evident from recent global estimates. Rumgay et al. estimated that 120,200 oral cancer cases diagnosed globally in 2022 were attributable to smokeless tobacco or areca nut consumption, corresponding to approximately 30.8% of all oral cancer cases (Rumgay et al., 2024). The burden was particularly high in south-central Asia. These findings demonstrate why population-level risk prediction should include detailed information about smokeless tobacco and areca nut exposure rather than relying only on smoking status. Alcohol is another established risk factor and the relationship between alcohol and oral cancer is influenced by dose and duration. Combined tobacco and alcohol exposure is especially important because the two exposures can interact in carcinogenesis. A digital risk assessment tool can capture smoking status, tobacco product type, and cessation history. Such information can be updated over time and it differs from a conventional risk assessment conducted during a single clinical consultation.

Age is also associated with oral cancer risk and the incidence generally increases with age. However, oral cancer can occur in younger adults when strong behavioral or genetic factors are present. A population model should therefore treat age as a continuous risk variable rather than defining oral cancer as a disease restricted to older adults. Oral hygiene, access to dental care, and previous oral lesions may also contribute to risk. These factors can be particularly important in community-based screening because populations with similar tobacco exposure may still experience different outcomes because of differences in healthcare access and social conditions. OPMDs represent another important component of population risk. The presence of an OPMD does not mean that cancer will inevitably develop. The probability of malignant transformation varies between lesion types and individual patients. Histological dysplasia is an important predictor, but clinical factors also contribute. Lesion size, anatomical

site, and previous lesion history can influence the risk profile (Warnakulasuriya et al., 2021) This heterogeneity creates a situation in which simple classification is insufficient. A patient with a small stable lesion and no major risk exposure may require a different follow-up strategy from a patient with extensive dysplasia and continued tobacco exposure. Machine learning is well suited to this type of multidimensional problem. A model can combine multiple independent variables and generate an estimated probability of progression. Uppal et al. systematically reviewed machine learning approaches for predicting malignant transformation of OPMDs and identified 15 eligible studies (Uppal et al., 2024). Reported overall prediction accuracy ranged from 74% to 95.4%. Some individual models achieved sensitivity or specificity of 100%. The review concluded that machine learning could incorporate multiple variables to estimate malignant potential, although performance remained sensitive to dataset characteristics. The findings from this systematic review are important because they demonstrate both the potential and limitation of AI-based risk prediction. High accuracy values indicate that computational models can identify useful patterns. However, variation between datasets means that a model developed in one population may not perform similarly elsewhere. The problem is particularly relevant to oral cancer because risk factors differ geographically. A model developed in a population with high areca nut use may identify different predictors from a model developed in a population where cigarette smoking and alcohol are dominant exposures. Digital epidemiology can address some of these challenges by creating larger and more diverse datasets. Digital epidemiology offers the possibility of updating risk when new information becomes available. A population-based system could consequently classify individuals according to their current risk rather than applying a fixed label. Low-risk individuals could remain within routine oral health services. Intermediate-risk individuals could receive periodic monitoring. High-risk individuals could receive more intensive clinical evaluation. Individuals with suspicious lesions could proceed to AI-assisted image assessment and specialist referral. This approach may allow limited specialist resources to be directed toward individuals with greater predicted need.

3. Digital Data Sources for Population-Based Oral Cancer Prediction

Electronic health records represent one of the most useful sources for population-based prediction because they contain information accumulated during routine healthcare. Longitudinal records are particularly valuable because oral cancer develops over time. A single consultation may capture only part of the risk profile. Repeated records can provide information about changing exposures and persistent lesions. However, electronic health records also contain missing data and inconsistent documentation. Tobacco and alcohol exposure may not be recorded in sufficient detail. Different hospitals may use different coding systems. Standardization is therefore necessary before data from multiple institutions can be combined. Clinical photographs represent another major source of digital information. The development of deep learning has made image-based oral cancer detection one of the most active areas of research. Fu et al. developed a deep learning algorithm using 44,409 clinical photographs collected from 11 hospitals (Fu et al., 2020). Internal validation produced an AUC of 0.983 with sensitivity of 94.9% and specificity of 88.7%. External validation produced an AUC of 0.935 with sensitivity of 89.6% and specificity of 80.6%. The model also demonstrated strong performance for early-stage OSCC. The study showed that ordinary clinical photographs can contain sufficient visual information for deep learning systems to identify patterns associated with oral cancer. The difference between internal and external validation performance was also informative because it demonstrated that model performance can decrease when the algorithm is applied to data from a different setting. Vision transformer approaches have also been investigated. Flugge et al. evaluated a Swin Transformer using 1,406 clinical photographs and reported an AUC of approximately 0.99 for OSCC detection (Flügge et al., 2023). Transformer-based models may provide advantages in capturing relationships across different regions of an image. However, improved architecture does not automatically eliminate problems related to dataset bias. A model can still learn characteristics associated with a particular camera, institution, or patient population. Multicenter research is therefore particularly important. Ye et al. evaluated a YOLOX-based deep learning model using 1,419 oral photographs obtained through a multicenter study involving 24 participants from 14 centers (Ye et al., 2024). High performance was reported across many studies. However, histopathological AI has its own sources of variability. Differences in staining procedures, tissue preparation, and annotation can affect model performance. A model trained in one pathology laboratory may therefore require recalibration before use in another institution. Hyperspectral imaging represents another developing source of information. Conventional photographs capture information within the visible spectrum. Hyperspectral imaging captures additional wavelength-specific information that may reflect differences in tissue composition. Li et al. evaluated a deep learning-based hyperspectral approach for identifying metastatic lymph node tissue in OSCC using specimens from 45 patients (Li et al., 2025). The improved ResUNet model achieved an overall accuracy of 87.30% and an average accuracy of 85.46% for tissue classification. Cancerous regions were identified with an intersection-over-union value of 0.6253. The study was preliminary and involved a relatively small cohort. Nevertheless, it demonstrates the possibility of combining spectral information with

deep learning to identify tissue characteristics that may not be visible in conventional photographs. Smartphones provide an important bridge between digital data and population screening. A 2026 literature review identified 23 studies published between 2015 and 2024 involving smartphone imaging, mHealth systems, tele-dentistry, autofluorescence, and AI-supported oral cancer detection (Alm et al., 2026). Reported sensitivity ranged from 70% to 99%, while specificity ranged from 64% to 100%. Reported accuracy ranged from 81% to 97%. These values indicate considerable potential, but the broad ranges also show that performance depends strongly on the system and study population. Smartphone-based autofluorescence systems provide another approach. The mDOC platform has been evaluated as a point-of-care imaging system for primary and community healthcare settings (Durab et al., 2025). The technology combines conventional white-light imaging with autofluorescence information. Such systems may become more useful when integrated with AI. A multimodal platform could combine patient risk factors, white-light images, autofluorescence images, and clinical history. The resulting model could provide a more comprehensive assessment than a single photograph (Table 1 & Figure 1).

Table 1: Emerging Digital Health Technologies and Artificial Intelligence Models Applied in Oral Cancer Risk Prediction

Data type	AI approach	Application	Contribution to digital epidemiology	Reference
Oral photographs	Object detection	OSCC and leukoplakia detection	Automated identification and localization of suspicious lesions	(Kouketsu et al., 2024)
Raman spectroscopy	ML ensembles	Early oral cancer risk assessment	Incorporates molecular information into computational risk assessment	(Chaudhuri et al., 2023)
Smartphone oral images	Deep neural networks	Premalignant/malignant lesion classification	Potential low-cost screening in remote communities	(Liyanage et al., 2023)
Histopathology	Multiple CNN architectures	Tissue-level classification	Supports digital pathology-assisted screening	(Afify et al., 2023)
Microscopic tissue images	CNN + Random Forest	Oral tissue classification	Combines deep feature extraction with conventional ML	(Jhawar & Pal, 2025)
Clinical photographs	Hybrid ML	OSCC detection	Demonstrates automated image analysis using relatively small datasets	(Al Duhayyim et al., 2023; Di Fede et al., 2025)
Clinical photographs	Deep learning	OSCC detection/ classification	Supports automated photographic screening	(Alanazi et al., 2022)
Intraoral lesion data	ML/DL	Automated POC classification	Supports decentralization of oral cancer screening	(Ferro et al., 2022)

4. Artificial Intelligence for Oral Cancer and OPMD Detection

The earliest applications of machine learning in oral cancer relied mainly on conventional algorithms such as artificial neural networks, random forests, and logistic regression. These models were often based on manually selected clinical or pathological features. Alabi et al. systematically reviewed 41 studies of machine learning in OSCC and reported sensitivity values ranging from 0.70 to 1.00, specificity from 0.57 to 1.00, and accuracy from 63.4% to 100% (Alabi et al., 2021). The wide range illustrates the variability in datasets and study design. It also demonstrates that high accuracy cannot be interpreted without considering the patient population, reference standard, sample size, and validation strategy.

Deep learning subsequently changed the field because these models can learn complex features directly from images. Warin and Suebnukarn identified 54 deep learning studies in oral cancer (Warin & Suebnukarn, 2024). Fifty-one studies evaluated diagnosis while only three investigated prognosis. Histopathological images were the most common data type, followed by computed tomography and clinical oral photographs. Reported classification accuracy generally ranged between 85% and 100%. The review concluded that deep learning showed promising performance but also identified heterogeneity and limitations in study quality. More recent meta-analyses provide broader evidence. A 2024 systematic review and meta-analysis evaluated AI for detecting OPMDs and oral cancer using medical diagnostic imaging (Sahoo et al., 2024). Fifty-five studies were included in the qualitative synthesis and 18 were included in the quantitative analysis. The pooled sensitivity was approximately 0.87 and specificity was approximately 0.81. The pooled diagnostic performance demonstrated a high SROC area under the curve. These results suggest that AI can distinguish suspicious and non-suspicious lesions with considerable accuracy, but the performance should not be assumed to be equivalent across all clinical environments.

The 2026 meta-analysis provides a more recent estimate (Soleimani Sardou et al., 2026). Sixteen eligible studies were included. AI models showed pooled sensitivity of 0.90, specificity of 0.89, and accuracy of 0.89. Image-based models achieved pooled sensitivity of 0.94 and specificity of 0.93. However, sensitivity across individual studies ranged from 42% to 100% and specificity ranged from 63% to 100%. The very high heterogeneity indicates that the pooled values represent an overall research estimate rather than a guaranteed clinical performance. OPMD detection is particularly important because prevention becomes possible before invasive cancer develops. Horvath et al. evaluated AI-assisted identification of OPMDs and reported promising diagnostic performance in their systematic review and meta-analysis (Horvath et al., 2026). This evidence suggests that the future of oral cancer AI should extend beyond lesion classification. The important question is not simply whether a lesion looks malignant. The system should ideally estimate the probability of malignant transformation and help determine the appropriate intensity of surveillance. Such prediction requires longitudinal data. A cross-sectional image can identify current appearance but cannot fully represent future biological behavior.

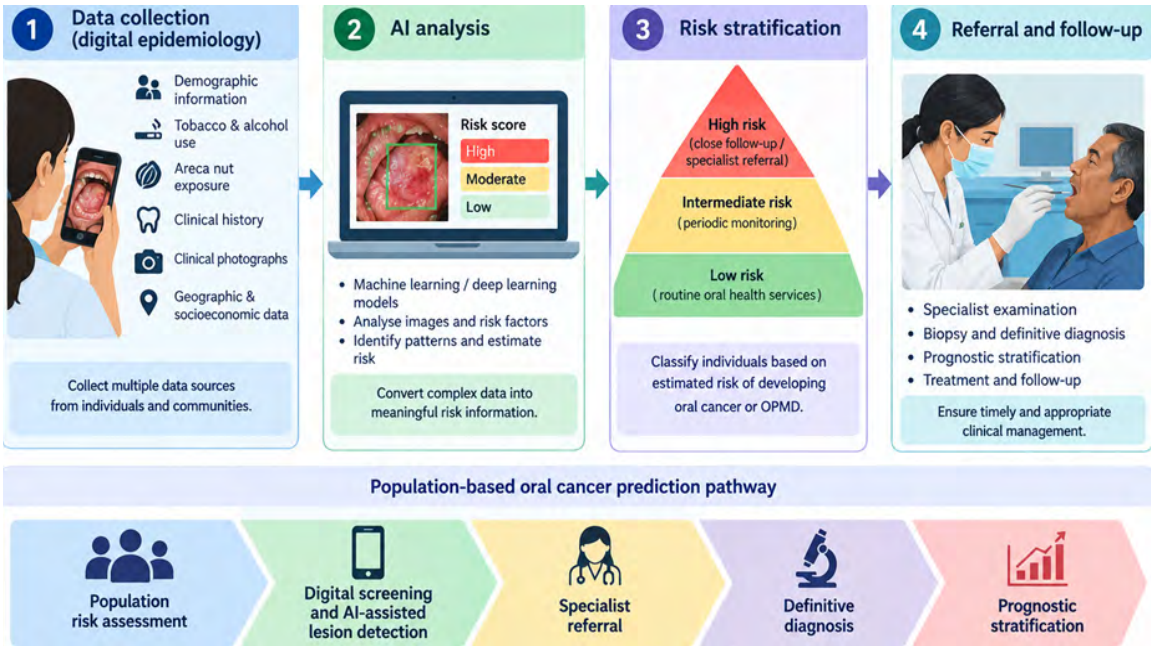


Figure 1: AI-enabled digital epidemiology framework for population-based oral cancer risk prediction, screening, risk stratification, and clinical referral.

5. AI-Based Risk Prediction and Prognostic Stratification

AI-based risk prediction differs from image-based diagnosis because the input variables can include behavioral, demographic, clinical, and pathological information. Alhazmi et al. provided an early example by developing an artificial neural network model for oral cancer risk prediction (Alhazmi et al., 2021). The study included 73 patients and 29 variables. Fifty-one patients had malignant disease and 22 had benign or premalignant conditions. Variables included age, clinical complaints, medical conditions, oral site, clinical appearance, shammah, khat, and tobacco exposure. Using 10-fold cross-validation, the neural network achieved a sensitivity of 85.71%, specificity of 60%, and accuracy of 78.95%. The study demonstrated the feasibility of combining clinical and behavioral factors for prediction. However, the small sample size limits the generalizability of the model. Larger community datasets are more appropriate for population-level risk prediction. Recent research from rural India illustrates this approach. Sundar et al. evaluated explainable machine learning for OPMD risk classification in a community-based cross-sectional study involving 3,700 adults from 100 rural clusters in Hassan District, Karnataka (M et al., 2026). The investigators collected demographic information, tobacco and alcohol exposure, oral hygiene information, and oral examination findings. XGBoost achieved an AUC of 0.91 and an accuracy of 85.7%. Age, tobacco exposure, and poor oral health were consistently important predictors. Cluster analysis identified a subgroup with substantially higher OPMD risk. The study is important because the model was developed using community data rather than a specialist hospital cohort. Such studies provide a stronger foundation for population-level screening.

AI has also been applied to predict lymph node metastasis. Nodal involvement has major implications for treatment planning in OSCC. A SEER-based population study evaluated six machine learning algorithms for predicting nodal metastasis in patients with early T-classification OSCC (Kwak et al., 2021). AUC values ranged from 0.768 to 0.956. XGBoost achieved the highest AUC at 0.956. Tumor size was identified as the most important feature using permutation importance. This finding demonstrates the ability of machine

learning to identify patterns associated with clinically important outcomes. Such models may provide additional information when treatment decisions are being considered. Large cancer registries also provide opportunities for survival prediction. Okeagu et al. analysed SEER data from 1992 to 2020 to develop machine learning models for oral cancer prognostication (Okeagu et al., 2026). The dataset included 53,611 cases, with 39,904 complete cases containing 21 variables. Random forest, XGBoost, LASSO, and logistic regression were compared. Random forest produced an AUC of approximately 77.6% and accuracy of 71.4%. SHAP analysis indicated that localized disease stage was the most influential variable, followed by age. The study also showed that discrimination and calibration can behave differently. A model may classify patients reasonably well while still producing poorly calibrated probabilities. Another 2026 SEER-based study developed a prognostic model for OSCC using 8,927 patients (Zhang et al., 2026). LASSO, XGBoost, random forest, and support vector machine algorithms were evaluated. Age, race, sex, tumor grade, and TNM stage were among the important predictors. The final model achieved a C-index of 0.714 compared with 0.665 for conventional TNM staging, with external validation reported by the investigators. These findings suggest that machine learning can integrate demographic and tumor characteristics to improve individualized prognostic estimation. The importance of these studies extends beyond algorithm development. Conventional staging remains central to clinical management because it is clinically interpretable and supported by extensive evidence. AI models may provide additional information rather than replace staging systems. A future prognostic platform could combine TNM stage with patient characteristics, pathological features, treatment information, and longitudinal outcomes. Such a model could generate individualized survival estimates and potentially support treatment discussions.

6. Population-Based Screening and Public Health Applications

Smartphone based systems can allow trained personnel to capture images and risk information without requiring an oral medicine specialist to be physically present. Recent research on the RiskOCA platform provides an example of this approach. The system combined patient risk profiling, guided smartphone imaging, deep learning, and specialist review. The model was trained using 2,226 annotated images and tested in 1,242 adults aged 40 years or older in rural Thailand. Mean AI classification accuracy across three diagnostic categories was approximately 93.22% (Wongsapai et al., 2026). The involvement of village health volunteers demonstrates how AI can be incorporated into an existing community health structure rather than functioning as a separate technology. Smartphone autofluorescence represents another possible pathway. The mDOC system was evaluated in a clinical imaging study involving 111 patients and 223 anatomical sites (Durab et al., 2025). The system combined white-light and autofluorescence imaging and was designed to support referral decisions in primary and community healthcare settings. The potential value of such technology may increase when optical imaging is combined with AI. The AI system could first determine whether image quality is sufficient. It could then combine white-light and fluorescence features with clinical risk information. The most effective AI system may therefore be one that connects risk prediction with an intervention pathway. A high-risk individual could receive counselling. A persistent lesion could trigger referral. A suspicious image could receive specialist review. A confirmed cancer could enter a prognostic pathway. This creates a continuum between epidemiology and clinical care.

7. Future Perspective

The future of AI-based oral cancer prediction is likely to involve multimodal systems rather than single-input models. Clinical photographs can describe visible lesions and behavioral data can describe exposure. Combining these information sources could produce a more complete representation of oral cancer risk. Federated learning may help develop such models without requiring all patient data to be transferred to a central database. Multiple hospitals could train a shared model while retaining patient-level information locally. This approach could increase dataset diversity while addressing some data-sharing concerns. Prospective studies will be required to determine whether federated systems provide reliable performance across institutions. Dynamic risk prediction is another important direction. Oral cancer risk changes over time and tobacco exposure can stop or increase. Digital systems can update risk when new information becomes available. A future platform could therefore generate an individualized risk trajectory rather than a single static score.

Explainable risk scores could also improve clinical use. A system could report an estimated risk together with the main contributing variables. Such output would be more informative than a simple “high-risk” label. The clinician could review the factors and determine whether the prediction is clinically reasonable. This approach would also make it easier to identify potential errors. Future studies should prioritize large prospective multicenter cohorts. Data should be collected from different geographical regions and socioeconomic populations.. Researchers should report sensitivity, specificity, predictive values, calibration, discrimination, and clinical utility. The effect of AI on referral rates and diagnostic delay should also be evaluated. Cost-effectiveness should be included because a technically accurate model may still have limited public health value if implementation is too expensive. The research agenda should also move toward integrated clinical pathways. A future population-based platform could begin with a digital questionnaire. The system could estimate individual risk

based on demographic and behavioral information. High-risk individuals could undergo targeted examination. Smartphone imaging could identify suspicious lesions. AI could prioritize cases for specialist assessment and histopathology could provide definitive diagnosis. AI could then assist with prognostic prediction after cancer confirmation. Public health authorities could use aggregated information to monitor geographic and temporal changes in disease risk. Such a framework would represent a significant change from the current image-centered AI literature. Most existing studies ask whether AI can recognize cancer from an image. The more important population-health question is whether AI can help identify people at risk before advanced disease develops and improve the pathway from risk recognition to definitive diagnosis. The recent evidence on OPMD transformation prediction, community risk classification, smartphone screening, and registry-based prognostic modelling suggests that the necessary components are beginning to emerge.

8. Conclusion

In conclusion, artificial intelligence and digital epidemiology provide complementary approaches for improving oral cancer early detection. AI offers the analytical capacity to integrate complex datasets. Digital epidemiology provides access to population-level information. The combination could support a continuous pathway from behavioral risk assessment to lesion detection and prognostic stratification. Current evidence is encouraging, but most models remain retrospective and require stronger external and prospective validation. The future value of AI in oral cancer will therefore depend not only on algorithmic accuracy but also on population representativeness and measurable improvement in patient outcomes. A clinically responsible system should complement specialist assessment rather than replace it. The ultimate objective should be earlier identification of high-risk individuals and more efficient movement of those patients into appropriate treatment pathways.

Declarations

Ethics approval statement

Not applicable

Consent to participate

Not applicable

Consent to publish

Not applicable

Data Availability Statement

The data are available from the corresponding author upon reasonable request

Competing Interests

The authors declare that they have no conflict of interest

Funding

Not Applicable

Author contribution

W.A, J.B, J.W, &, L.P equally contributed in preparation of this manuscript. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

Not Applicable

Reference

1. Afify, H. M., Mohammed, K. K., & Ella Hassanien, A. (2023). Novel prediction model on OSCC histopathological images via deep transfer learning combined with Grad-CAM interpretation. Biomedical Signal Processing and Control, 83, 104704. https://doi.org/10.1016/j.bspc.2023.104704
2. Al Duhayyim, M., A. Malibari, A., Dhahbi, S., K. Nour, M., Al-Turaiki, I., Obayya, M., & Mohamed, A. (2023). Sailfish Optimization with Deep Learning Based Oral Cancer Classification Model. Computer Systems Science and Engineering, 45(1), 753–767. https://doi.org/10.32604/csse.2023.030556
3. 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. Artificial Intelligence in Medicine, 115, 102060. https://doi.org/10.1016/j.artmed.2021.102060
4. Alanazi, A. A., Khayyat, M. M., Khayyat, M. M., Elamin Elnaim, B. M., & Abdel-Khalek, S. (2022). Intelligent Deep Learning Enabled Oral Squamous Cell Carcinoma Detection and Classification Using Biomedical Images. Computational Intelligence and Neuroscience, 2022, 1–11. https://doi.org/10.1155/2022/764396
5. Alhazmi, A., Alhazmi, Y., Makrami, A., Masmali, A., Salawi, N., Masmali, K., & Patil, S. (2021). Application of artificial intelligence and machine learning for prediction of oral cancer risk. Journal of Oral Pathology & Medicine, 50(5), 444–450. https://doi.org/10.1111/jop.13157
6. Alm, H., Salo, T., & Elseragy, A. (2026). Artificial intelligence–driven smartphone imaging for early detection of oral cancer: How far have we come? A comprehensive literature review. Dentistry Review, 6(1), 100162. https://doi.org/10.1016/j.dentre.2026.100162
7. Chaudhuri, D., Ghosh, A., Raha, S., Chakraborty, A., Chatterjee, K., & Barui, A. (2023). Machine learning algorithm ensembles for early oral cancer risk assessment using Raman cyto-spectroscopy. Soft Computing, 27(19), 13861–13875. https://doi.org/10.1007/s00500-023-08995-z
8. Di Fede, O., La Mantia, G., Parola, M., Maniscalco, L., Matranga, D., Tozzo, P., Campisi, G., & Cimino, M. G. C. A. (2025). Automated

Detection of Oral Malignant Lesions Using Deep Learning: Scoping Review and Meta-Analysis. Oral Diseases, 31(4), 1054–1064. https://doi.org/10.1111/odi.15188

9. Durab, S., Mitbander, R., Coole, J., Kortum, A., Vohra, I., Carns, J., Schwarz, R., Varghese, I., Anderson, S., Badaoui, H., Anandasivam, L., Giese, R., Gillenwater, A., Richards-Kortum, R., & Vigneswaran, N. (2025). A clinical imaging study using a smartphone point-of-care autofluorescence device to facilitate the early detection of oral cancer in primary and community health clinical settings. Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology, 140(3), e93. https://doi.org/10.1016/j.oooo.2025.04.081
10. Ferro, A., Kotecha, S., & Fan, K. (2022). Machine learning in point-of-care automated classification of oral potentially malignant and malignant disorders: a systematic review and meta-analysis. Scientific Reports, 12(1), 13797. https://doi.org/10.1038/s41598-022-17489-1
11. Flügge, T., Gaudin, R., Sabatakakis, A., Tröltzsch, D., Heiland, M., van Nistelrooij, N., & Vinayahalingam, S. (2023). Detection of oral squamous cell carcinoma in clinical photographs using a vision transformer. Scientific Reports, 13(1), 2296. https://doi.org/10.1038/s41598-023-29204-9
12. 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., … Xiong, X. (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
13. Gupta, S., Jawanda, M. K., & Madhushankari, G. (2020). Current challenges and the diagnostic pitfalls in the grading of epithelial dysplasia in oral potentially malignant disorders: A review. Journal of Oral Biology and Craniofacial Research, 10(4), 788–799. https://doi.org/10.1016/j.jobcr.2020.09.005
14. Horvath, D., Fekete, A., Martinekova, P., Kiss-Dala, S., Abram, E., Marton, K., Zsembery, A., Hegyi, P., & Brody, A. (2026). Artificial Intelligence Helps Diagnose Oral Potentially Malignant Disorders: A Systematic Review and Meta-Analysis. JDR Clinical & Translational Research, 11(4), 532–544. https://doi.org/10.1177/23800844251403962
15. Jhawar, R., & Pal, A. (2025). Machine Learning and Deep Learning Approaches for Cancer Histopathology Image Classification. International Journal for Research in Applied Science and Engineering Technology, 13(11), 525–529. https://doi.org/10.22214/ijraset.2025.74927
16. Kouketsu, A., Doi, C., Tanaka, H., Araki, T., Nakayama, R., Toyooka, T., Hiyama, S., Iikubo, M., Osaka, K., Sasaki, K., Nagai, H., Sugiura, T., Yamauchi, K., Kuroda, K., Yanagisawa, Y., Miyashita, H., Kajita, T., Iwama, R., Kurobane, T., & Takahashi, T. (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
17. Kwak, M. S., Eun, Y., Lee, J., & Lee, Y. C. (2021). Development of a machine learning model for the prediction of nodal metastasis in early T classification oral squamous cell carcinoma: SEER -based population study. Head & Neck, 43(8), 2316–2324. https://doi.org/10.1002/hed.26700
18. Li, Q., Zhang, X., Zhang, J., Huang, H., Li, L., Guo, C., Li, W., & Guo, Y. (2025). Deep learning-based hyperspectral technique identifies metastatic lymph nodes in oral squamous cell carcinoma—A pilot study. Oral Diseases, 31(2), 417–425. https://doi.org/10.1111/odi.15067
19. Liyanage, V., Tao, M., Park, J. S., Wang, K. N., & Azimi, S. (2023). Malignant and non-malignant oral lesions classification and diagnosis with deep neural networks. Journal of Dentistry, 137, 104657. https://doi.org/10.1016/j.jdent.2023.104657
20. M, S., M, S., Khot, P. B., & S, M. S. F. (2026). Risk classification of oral potentially malignant disorders (OPMDs) using explainable machine learning through a community-based cross-sectional study in rural India. BMC Public Health, 26(1), 1972. https://doi.org/10.1186/s12889-026-27283-7
21. Okeagu, P., Murukutla, S., Iwuchukwu, A., & Ogwo, C. (2026). A Machine Learning Approach to Prognostication in Oral Cancer: Analysis of the Surveillance, Epidemiology, and End Results Database. Cancer Informatics, 25. https://doi.org/10.1177/11769351261442875
22. Rumgay, H., Nethan, S. T., Shah, R., Vignat, J., Ayo-Yusuf, O., Chaturvedi, P., Guerra, E. N. S., Gupta, P. C., Gupta, R., Liu, S., Magnusson, C., Parascandola, M., Paulino, Y. C., Rezaei, N., Siddiqi, K., Warnakulasuriya, S., Lauby-Secretan, B., & Soerjomataram, I. (2024). Global burden of oral cancer in 2022 attributable to smokeless tobacco and areca nut consumption: a population attributable fraction analysis. The Lancet Oncology, 25(11), 1413–1423. https://doi.org/10.1016/S1470-2045(24)00458-3
23. Sahoo, R. K., Sahoo, K. C., Dash, G. C., Kumar, G., Baliarsingh, S. K., Panda, B., & Pati, S. (2024). Diagnostic performance of artificial intelligence in detecting oral potentially malignant disorders and oral cancer using medical diagnostic imaging: a systematic review and meta-analysis. Frontiers in Oral Health, 5. https://doi.org/10.3389/froh.2024.1494867
24. Soleimani Sardou, S., Rezvaninejad, R., Rezvaninejad, F. S., & Nekouei, A. H. (2026). Artificial intelligence for oral cancer diagnosis: a systematic review and meta-analysis of image-based and non-imaging models. BMC Cancer, 26(1), 868. https://doi.org/10.1186/s12885-026-16154-4

25. 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 Journal for Clinicians, 71(3), 209–249. https://doi.org/10.3322/caac.21660
26. Uppal, S., Kumar Shrivastava, P., Khan, A., Sharma, A., & Kumar Shrivastav, A. (2024). Machine learning methods in predicting the risk of malignant transformation of oral potentially malignant disorders: A systematic review. International Journal of Medical Informatics, 186, 105421. https://doi.org/10.1016/j.ijmedinf.2024.105421
27. 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
28. Warnakulasuriya, S., Kujan, O., Aguirre-Urizar, J. M., Bagan, J. V., González-Moles, M. Á., Kerr, A. R., Lodi, G., Mello, F. W., Monteiro, L., Ogden, G. R., Sloan, P., & Johnson, N. W. (2021a). Oral potentially malignant disorders: A consensus report from an international seminar on nomenclature and classification, convened by the WHO Collaborating Centre for Oral Cancer. Oral Diseases, 27(8), 1862–1880. https://doi.org/10.1111/odi.13704
29. Wongsapai, M., Wudtijureepun, K., Suthachai, T., Tamdee, D., Ittichaicharoen, J., Wuttisarnwattana, P., & Suebnukarn, S. (2026). A Community-Based Usability Study of an AI-Enabled Oral Cancer Screening App Operated by Village Health Volunteers: Mixed Methods Study. JMIR MHealth andUHealth, 14, e83738–e83738. https://doi.org/10.2196/83738
30. Ye, Y.-J., Han, Y., Liu, Y., Guo, Z.-L., & Huang, M.-W. (2024). Utilizing deep learning for automated detection of oral lesions: A multicenter study. Oral Oncology, 155, 106873. https://doi.org/10.1016/j.oraloncology.2024.106873
31. Zhang, Y., Liu, H., Liao, Y., Su, Z., Song, L., Ma, Z., Pan, Z. qiang, Chen, Y., Xia, M., Li, J., & Jiang, L. (2026). Machine learning based prognostic model for oral squamous cell carcinoma using SEER data and external validation. IScience, 29(5), 115634. https://doi.org/10.1016/j.isci.2026.115634