

Systematic Review

Performance of Machine Learning Models for Prognosis Prediction in Oral Cavity Squamous Cell Carcinoma: A Systematic Review

Sammy Y. Gao, Jonathan M. Hughes, Shaun A. Nguyen *, Bryce S. McCaulay and Jason G. Newman

Department of Otolaryngology—Head and Neck Surgery, Medical University of South Carolina, Charleston, SC 29425, USA; gaos@musc.edu (S.Y.G.); hughejon@musc.edu (J.M.H.); macaulab@musc.edu (B.S.M.); newmajas@musc.edu (J.G.N.)

* Correspondence: nguyensh@musc.edu

Simple Summary

Oral cavity squamous cell carcinoma is a common type of oral cancer, and predicting which patients are most likely to experience recurrence, disease progression, or reduced survival remains challenging. In recent years, machine learning has been increasingly used to analyze large amounts of clinical, imaging, pathology, and molecular data to improve prognostic predictions. This systematic review examines published studies that developed machine learning models to predict outcomes in patients with oral cavity squamous cell carcinoma. Across 40 studies involving more than 105,000 patients, many models showed good ability to predict survival and recurrence. However, differences in study design, reporting quality, and validation methods limit confidence in their widespread clinical use. These findings highlight the promise of machine learning for personalized cancer care while emphasizing the need for larger, well-validated studies before routine implementation in clinical practice.

Abstract

Background/Objectives: Machine learning (ML) models have increasingly been applied to prognostic prediction in oral cavity squamous cell carcinoma (OCSCC), though their performance and methodological quality remain variably reported. This systematic review evaluated contemporary ML-based prognostic models for clinically relevant OCSCC outcomes, with emphasis on independently validated studies. **Methods:** A systematic review was conducted according to PRISMA guidelines using PubMed, Scopus, Cochrane Library, and CINAHL databases through 1 December 2025. Studies evaluating ML or artificial intelligence prognostic models in adult OCSCC patients were included. Outcomes included overall survival, recurrence, disease-free survival, recurrence-free survival, disease-specific survival, cancer-specific survival, progression, and nodal metastasis. Data extraction and risk-of-bias assessment using PROBAST + AI were performed independently by reviewers. **Results:** Forty studies comprising 105,619 patients met inclusion criteria. ML architectures included random forests, support vector machines, gradient boosting methods, neural networks, and deep learning frameworks. Most models incorporated clinical and pathologic variables, while many integrated radiologic, immunologic, or genomic features. For overall survival prediction, independently validated models generally demonstrated AUCs between 0.80 and 0.90. Recurrence prediction models similarly showed favorable discrimination, with most externally validated studies reporting acceptable predictive performance. Additional

Academic Editors: Maxim V. Berezovski and Ognjen Arandjelović

Received: 1 June 2026

Revised: 30 June 2026

Accepted: 13 July 2026

Published: 14 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

prognostic endpoints including disease-free survival, progression, and nodal metastasis demonstrated AUCs ranging from 0.70 to 0.90. Common methodological limitations included retrospective design, small sample size, inadequate external validation, and risk of overfitting. **Conclusions:** ML-based prognostic models in OCSCC demonstrate generally favorable predictive performance across survival and recurrence outcomes. However, substantial heterogeneity in methodology and limited external validation continue to restrict clinical implementation. Future work should prioritize prospective multicenter validation, standardized reporting, and reproducible modeling frameworks.

Keywords: head and neck cancer; oral cavity; squamous cell carcinoma; prognosis; machine learning; overall survival

1. Introduction

Oral cavity squamous cell carcinoma (OCSCC) is a prevalent malignancy worldwide and remains associated with substantial morbidity and mortality despite advances in surgical and oncologic management [1–3]. Prognosis in OCSCC is strongly influenced by factors including locoregional recurrence, nodal metastasis, and disease progression, making accurate prognostic stratification clinically important for treatment selection, surveillance planning, and patient counseling. Historically, prognostic assessment in OCSCC has relied on conventional statistical approaches such as logistic regression, Cox proportional hazards modeling, nomograms, and clinicopathologic staging systems [4,5]. While these methods have demonstrated utility, they are often limited in their ability to account for complex nonlinear interactions among clinical, radiologic, pathologic, and molecular variables [6].

Machine learning (ML) methods offer an alternative framework capable of identifying high-dimensional relationships within large datasets while reducing reliance on predefined assumptions regarding predictor interactions. Unlike traditional statistical models, ML approaches can integrate multimodal data sources and adaptively optimize predictive performance [7,8]. Over the past decade, increasing numbers of studies have evaluated ML-based prognostic models in OCSCC for outcomes including overall survival, recurrence, disease-specific survival, and progression. Reported performance metrics have generally demonstrated favorable discriminatory ability, generating interest in the potential role of ML-assisted prognostic stratification in head and neck oncology.

However, important methodological concerns remain. Many published studies utilize retrospective single-center datasets, limited sample sizes, or internal validation approaches that may overestimate true model performance [9]. In addition, substantial heterogeneity exists regarding predictor selection, model architecture, outcome definitions, and reporting standards [10]. Prior reviews examining artificial intelligence in head and neck oncology have broadly summarized these technologies, but few have focused specifically on prognosis prediction in OCSCC using contemporary ML methodologies and independent validation cohorts [9].

Therefore, the objective of the present systematic review was to evaluate the performance of ML models developed for prognosis prediction in OCSCC, with specific emphasis on independently validated prognostic models and clinically relevant outcomes including overall survival and recurrence. This review additionally aimed to characterize model architectures, input modalities, validation strategies, and methodological quality across the existing literature.

2. Materials and Methods

2.1. Literature Search

This systematic review was conducted in accordance with Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (File S1) [11,12]. This systematic review was registered with the Open Science Framework (OSF; Registration ID: JP3FX). A comprehensive literature search of PubMed, Scopus, Cochrane Library, and CINAHL databases was performed from database inception through 1 December 2025. Search strategies incorporated combinations of Medical Subject Headings (MeSH), keywords, and Boolean operators related to oral cavity squamous cell carcinoma, prognosis, survival, recurrence, machine learning, artificial intelligence, deep learning, and predictive modeling. Full search strategies are provided in Supplemental Table S1. Retrieved studies were imported into Covidence systematic review software for screening and management (Veritas Health Innovation Ltd., Australia). Articles published online ahead of print and indexed prior to the search end date were considered eligible for inclusion regardless of their subsequent assignment to a print publication year. The search was limited to English-language publications, and grey literature was not searched.

2.2. Eligibility Criteria

Studies were eligible for inclusion if they: (1) included adult patients with OSCC; (2) evaluated a machine learning or artificial intelligence-based prognostic prediction model; (3) assessed outcomes related to prognosis, including overall survival (OS), recurrence, disease-free survival (DFS), recurrence-free survival (RFS), progression, disease-specific survival (DSS), cancer-specific survival (CSS), or nodal metastasis; and (4) reported performance metrics including area under the receiver operating characteristic curve (AUC), concordance index (C-index), accuracy, sensitivity, or specificity.

Studies utilizing exclusively traditional statistical models without a machine learning component were excluded. Reviews, editorials, conference abstracts without extractable data, and studies not isolating oral cavity primary tumors were also excluded. Independent validation was defined as model evaluation using data not utilized during model training, including separate institutional cohorts, external validation datasets, or predefined held-out test cohorts. Internal validation referred to split-sample validation, cross-validation, or bootstrapping performed within the original development dataset.

2.3. Data Extraction

Two reviewers (S.Y.G. and B.S.M.) independently extracted study characteristics, patient demographics, tumor characteristics, model inputs, model algorithms, outcome measures, validation methods, and reported performance metrics. Disagreements were resolved by a third reviewer (S.A.N.). Extracted variables included study country, data source, model architecture, input modality, and prognostic endpoints. Input domains were categorized as clinical, pathologic, radiologic, immunologic, genomic/molecular, or multimodal combinations. Machine learning architectures included tree-based models, support vector machines (SVMs), ensemble learning methods, neural networks, deep learning approaches, and survival-based machine learning frameworks. For studies which evaluated multiple algorithms or varied model inputs, the highest AUC for each outcome was extracted.

2.4. Risk of Bias Assessment

Risk of bias was assessed using the PROBAST + AI tool for prediction model development and evaluation studies by S.Y.G. and B.S.M. with disagreements resolved by a third reviewer (S.A.N.) [13]. Domains included participant selection, predictor

assessment, outcome determination, and analytical methodology. Particular attention was given to cohort size, handling of missing data, overfitting risk, validation methodology, and transparency of model reporting. Studies were categorized as low, high, or unclear risk of bias.

2.5. Statistical Analysis

Study, patient, and clinicopathologic characteristics were summarized descriptively because of substantial heterogeneity and inconsistent reporting across included studies. Continuous variables are presented as the median reported study-level value and range, while categorical variables are presented as the median reported study-level proportion and range among studies reporting that characteristic. Performance metrics for machine learning models, including AUC and C-index, were summarized descriptively in tabular and graphical form rather than quantitatively pooled because of substantial heterogeneity in model architecture, predictor selection, outcome definitions, validation methodology, and incomplete reporting of variance estimates. Definitions of these performance metrics are summarized in Supplementary Figure S1. As a qualitative sensitivity assessment, reported model performance was additionally examined according to validation strategy (independent versus internal validation) and overall PROBAST + AI risk-of-bias classification to evaluate the robustness of conclusions with respect to methodological quality. All statistics were performed in R version 4.6.0 (R Core Team).

3. Results

3.1. Included Studies

A total of 40 studies met inclusion criteria, encompassing 105,619 patients with OCSCC (Figure 1; Supplemental Table S2). Risk of bias assessed using PROBAST + AI was generally low in the participant, predictor, and outcome domains but more variable in the analysis domain, where concerns frequently related to small sample size, limited outcome events, incomplete reporting or handling of missing data, and lack of robust validation (Supplemental Tables S3 and S4). Overall, studies utilizing larger multicenter or registry-based cohorts with independent validation exhibited lower overall risk of bias and more plausible performance estimates, whereas small, single-center studies employing complex deep learning or multimodal architectures were more likely to report near-perfect discrimination suggestive of optimism bias.

Most investigations employed retrospective cohort designs using registry-based datasets, single-institution cohorts, or multi-institutional databases drawn from North America, Europe, and Asia, including the United States, China, Taiwan, Japan, South Korea, and several European countries. Larger sample sizes were typically observed in registry and multicenter datasets, whereas radiomic and genomic investigations more commonly relied on smaller institutional cohorts.

Across included studies, the median reported patient age was 61 years (range, 50.0–65.3 years; Table 1), and the median proportion of male patients was 68.6% (range, 47.8–92.6%). The tongue was the most frequently represented primary tumor subsite, with a median study-level proportion of 67.6% (range, 33.9–100.0%) among studies reporting tumor location. Adverse histopathologic features, including perineural invasion, lymphovascular invasion, bone invasion, and extranodal extension, were variably reported across studies (Table 1). Most prognostic models incorporated clinical and pathologic covariates, with an increasing proportion of studies integrating radiologic, immunologic, genomic, and multimodal inputs in more recent publications.

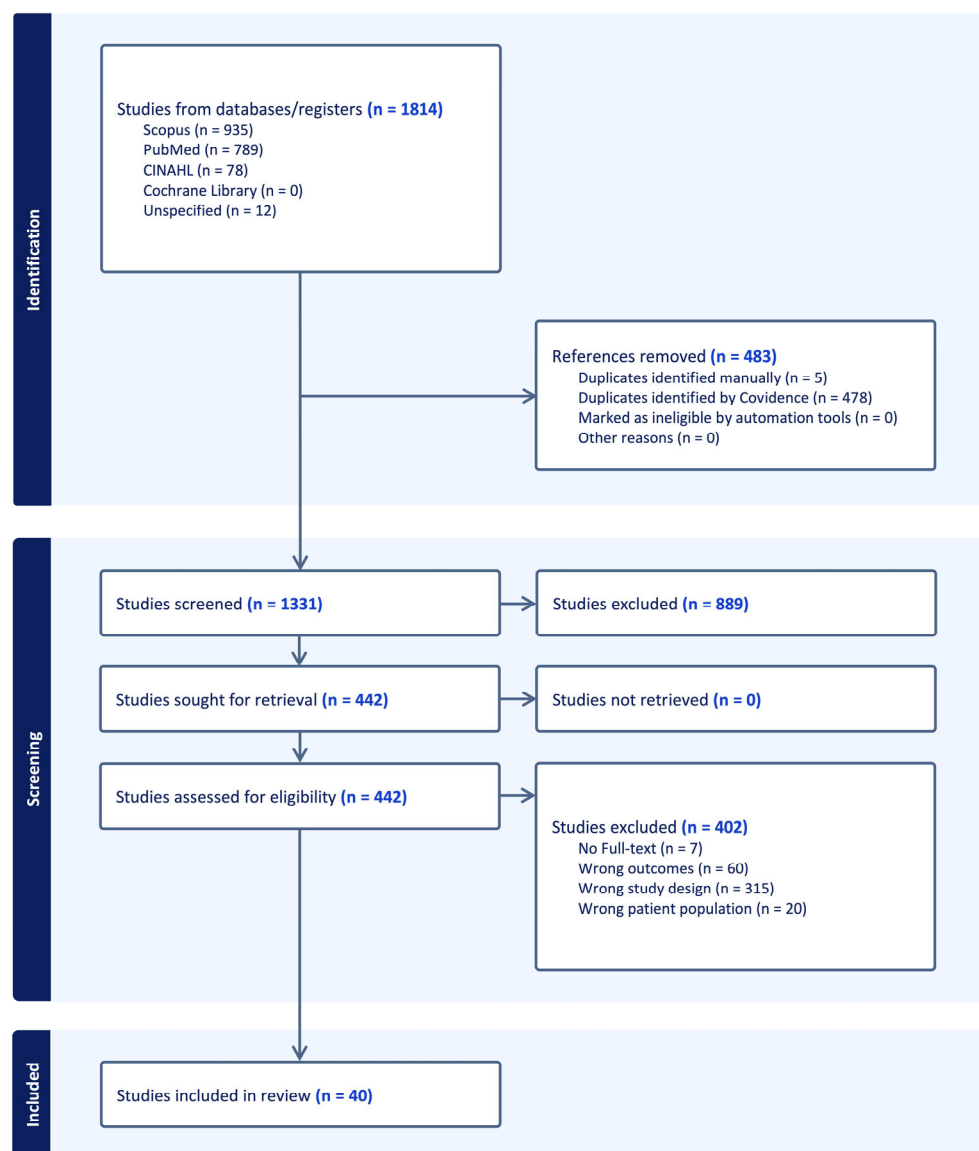


Figure 1. PRISMA Flow Diagram. Copyright statement: this PRISMA diagram contains public sector information licensed under the Open Government License v3.0 [11].

A broad range of machine learning architectures were evaluated, including random forests, decision trees, support vector machines (SVMs), gradient boosting methods, neural networks, deep learning frameworks, survival forests, and ensemble learning models with traditional regression-based or nomogram-based approaches.

Table 1. Demographics.

Characteristic	Number of Studies	Number of Patients	Median	Range
Age	17	105,619	61.1	50–65.3
Male sex	28	73,142	68.6%	47.8–92.6%
White ethnicity	6	30,767	86.6%	78.5–90.4%
Black ethnicity	9	64,710	7.1%	3.5–12.1%
<i>Site</i>				
Tongue	24	72,311	67.6%	33.9–100%
Gingiva/gum	12	53,249	19.6%	6.3–55.7%
Buccal mucosa	10	43,730	16.9%	4–64.1%
Floor of mouth	12	46,382	10.3%	0.6–42.2%

Retromolar	8	45,282	3.2%	2.6–13.7%
Hard palate	10	52,984	2.3%	1.6–15.7%
Lips	7	49,467	11.5%	0.5–26.8%
PNI	12	3716	23.3%	8.2–73.8%
LVI	11	3405	17.2%	4–73.8%
BNI	4	1210	22.6%	9.6–31.7%
ECE	10	3095	16.3%	7.6–76.2%
<i>TNM Staging</i>				
Tumor (T)				
1	21	28,712	35.8%	0–98.2%
2	21	28,486	33.0%	0–62.7%
3	20	28,315	14.9%	0.6–100%
4	17	27,858	24.10%	0–55.9%
Nodal Status (N)				
0	18	20,833	59.30%	26.1–100%
1	18	28,020	15.60%	7.7–44.9%
2	19	28,070	23.10%	0–76%
3	18	28,020	2.20%	0–16.9%
Metastasis (M)				
0	8	18,520	98.60%	35.7–100%
1	10	25,539	0.40%	0–4.9%

Values are summarized at the study level rather than pooled across all patients. For percentage variables, the median represents the median study-level proportion among studies reporting that characteristic, and the range represents the minimum and maximum reported study-level proportions. For example, the median tongue primary site proportion of 67.6% indicates that among the 24 studies reporting tongue involvement, the median study contained 67.6% tongue tumors. Because reporting varied across studies and many cohorts were restricted to specific oral cavity subsites, percentages are not derived from a common denominator and should not be summed across categories. Similarly, TNM stage percentages represent study-level medians and should not be expected to sum to 100%.

3.2. Overall Survival Prediction

Overall survival (OS) was the most frequently evaluated endpoint, with model performance reported in 20 studies (Table 2, Figure 2). Reported AUC values for OS prediction ranged from 0.60 to 1.00, while C-indices typically varied between 0.70 and 1.00. Among models with independent validation, most AUCs clustered between 0.80 and 0.90, indicating acceptable-to-strong discriminatory ability for OS. Few studies reported calibration measures, decision-curve analyses, or direct comparisons with established prognostic systems.

Table 2. Performance of Machine Learning Models for Overall Survival Prediction.

Study	Input Type	Model Algorithm	C-Index	AUC	Accuracy (%)	Sensitivity (%)	Specificity (%)
<i>Independent Validation</i>							
Somyanonthanakul 2024 [14]	Clinical, Pathologic	Fuzzy Deep Learning		1.00	97.0	90.0	97.0
Lee 2025 [15]	Clinical, Pathologic, Radiologic	Random Forest		0.89		86.0	73.8
Alabi 2026 [16]	Clinical	Voting ensemble		0.79	70.4	68.0	73.0
Karadaghy 2019 [6]	Clinical, Pathologic	Decision Forest		0.80	71.0	68.0	
Mes 2020 [17]	Radiologic, Clinical	Combined radiomic + clinical Cox model		0.82			
Han 2025 [18]	Genomic/Molecular	Elastic Net		0.74			
Deepali 2025 [19]	Clinical, Genomic/Molecular	DeepOmicsSurv	1.00				

Diao 2021 [20]	Immunologic	SVM	0.71			
Kim 2019 [21]	Pathologic	DeepSurv	0.78			
Peng 2022 [22]	Clinical	Random forest	0.80			
<i>Internal Validation</i>						
Alabi 2023 [23]	Clinical	Extreme random tree	0.91	86.0	80.0	90.0
Alabi 2021 [24]	Clinical	Boosted Decision Tree	0.90	83.1	85.0	
Li 2023 [25]	Clinical, Immunologic	Light Gradient Boosting Machine	0.86			
Huang 2020 [26]	Genomic/Molecular, Immunologic	Random forest + SVM	0.85			
Lin 2025 [27]	Clinical	DeepSurv-based deep neural network	0.78	0.75	93.1	56.9
Tan 2022 [28]	Clinical	Random Forest classifier	0.61	62.0	59.0	
Adeoye 2022 [29]	Clinical, Pathologic	DeepSurv	0.77			
Nezamabadi Farahani 2024 [30]	Genomic/Molecular	Particle Swarm Optimization + SVM	0.97			
Rosado 2013 [31]	Clinical, Genomic/Molecular	SVM		98.6		
Vollmer 2024 [7]	Clinical, Pathologic, Radiologic, Genomic/Molecular	Random Survival Forest	0.83			

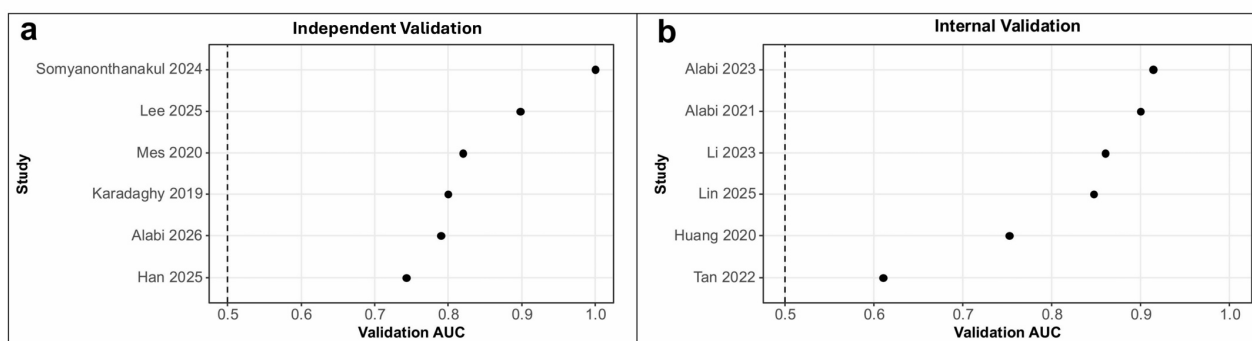


Figure 2 (a) [6,14,15,17,23,25], (b) [23–28]. Independent and Internal Validation AUCs of Machine Learning Models for Overall Survival Prediction. Dot plots include only studies reporting AUC values for the corresponding outcome. Therefore, the number of studies displayed in the figures may differ from the total number of studies included in the outcome-specific tables.

Commonly used architectures for OS prediction included ensemble learning approaches, random forests, SVMs, and deep neural network-based survival models. Input features most frequently combined clinical and pathologic variables, although several independently validated models additionally integrated radiologic imaging features, immune signatures, or genomic profiles. Notably, independently validated OS models often demonstrated performance metrics comparable to internally validated algorithms, whereas extreme near-perfect AUC values were largely confined to smaller, internally validated datasets, consistent with overfitting and optimism bias.

3.3. Recurrence Prediction

Recurrence outcomes, including locoregional recurrence and composite recurrence endpoints, were evaluated in 11 studies (Table 3, Figure 3). Across these models, reported AUC values ranged from 0.60 to 1.00, with most independently validated recurrence models demonstrating AUC values within acceptable prognostic discriminatory thresholds (approximately ≥ 0.70). Frequently employed model architectures for recurrence prediction included support vector machines, random forests, multilayer perceptrons, gradient boosting methods, and XGBoost-based algorithms. Clinical and pathologic variables formed the core predictor set, while several studies augmented these features with radiologic radiomics or immunologic markers. Independently validated recurrence models generally maintained performance metrics similarly to internally

validated counterparts, suggesting potential generalizability despite variation in predictor selection and modeling strategies.

Table 3. Performance of Machine Learning Models for Recurrence Prediction.

Study	Input Type	Model Algorithm	C-Index	AUC	Accuracy (%)	Sensitivity (%)	Specificity (%)
<i>Independent Validation</i>							
Adachi 2024 [32]	Digital Pathology, Pathologic	CLAM + SVM		0.99	94.9		
Cai 2023 [33]	Clinical, Pathologic, Immunologic	Multilayer perceptron		0.89	83.5	74.0	
Chen 2025 [34]	Genomic/Molecular, Clinical	Random Forest		0.85			
Vidiri 2025 [35]	Clinical, Pathologic	Support Vector Machine		0.79	67.0	67.0	50.0
Fatapour 2023 [36]	Clinical	Gradient Boosting Machine		0.74	80.0	82.8	
Huang 2025 [37]	Clinical, Pathologic	XGBoost		0.58	57.9	57.1	
<i>Internal Validation</i>							
Alabi 2020 [38]	Clinical, Pathologic	Boosted Decision Tree			81.0	79.0	83.0
Bourdillon 2023 [39]	Clinical, Pathologic	Decision tree classifier			80.8	80.8	
Alabi 2019 [40]	Clinical, Pathologic	Artificial neural network		0.97	88.2	71.2	98.9
Salehi 2024 [41]	Clinical, Pathologic, Radiologic	Random Forest		0.73	71.2	58.3	78.6
Pfaehler 2025 [42]	Radiologic	Random Forest		0.63	70.0	76.0	50.0

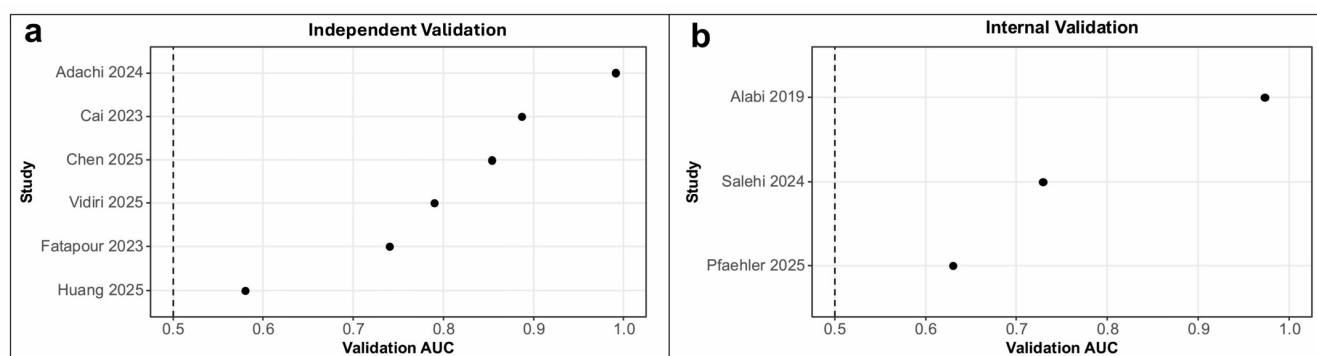


Figure 3 (a) [33–37], (b) [38,41,42]. Independent and Internal Validation AUCs of Machine Learning Models for Recurrence Prediction. Dot plots include only studies reporting AUC values for the corresponding outcome. Therefore, the number of studies displayed in the figures may differ from the total number of studies included in the outcome-specific tables.

3.4. Additional Prognostic Outcomes

Several studies developed machine learning models for additional prognostic endpoints, including (DFS), recurrence-free survival (RFS), disease-specific survival (DSS), cancer-specific survival (CSS), progression, and nodal metastasis (Table 4). Across these outcomes, AUC values typically ranged from 0.70 and 0.90, indicating moderate-to-good discriminatory ability for most models. These analyses utilized diverse algorithms such as random survival forests, decision trees, naïve Bayes classifiers, k-nearest neighbors, deep learning architectures, and radiomic convolutional neural networks. Predictor sets varied widely but generally comprised combinations of clinical staging variables, histopathologic features, imaging-derived radiomics, immunologic signatures, and genomic or transcriptomic markers. Multimodal models often yielded the highest

reported performance metrics, although direct comparison across studies was limited by differences in outcome definitions, feature engineering pipelines, and validation strategies.

Table 4. Performance of Machine Learning Models for Prediction of Metrics of Prognosis.

Study	Input Type	Model Algorithm	Validation Type	C-Index	AUC	Accuracy (%)	Sensitivity (%)	Specificity (%)
<i>DFS/RFS</i>								
Alkhadar 2021 [43]	Clinical, Pathologic	Decision tree classifier	Independent		0.77	76.0	81.0	72.0
Mes 2020 [17]	Radiologic, Clinical	Combined radiomic + clinical Cox model	Independent		0.76			
Diao 2021 [20]	Immunologic	SVM	Independent	0.68				
Fujima 2020 [44]	Radiologic	ResNet-101 deep learning classifier	Independent			80.0	80.0	80.0
Tseng 2015 [45]	Clinical, Pathologic	Decision tree	Internal			95.8		
<i>Progression</i>								
Mei 2025 [46]	Clinical, Pathologic, Immunologic	K Neighbors Classifier	Independent		0.69	74.5	47.1	80.7
Chu 2020 [47]	Clinical, Pathologic	K-nearest neighbours	Internal		0.71	69.4	35.1	85.6
<i>Nodal Metastasis</i>								
Csúry 2024 [48]	Pathologic, Clinical	Naïve Bayes	Independent		0.83			
Pfaehler 2025 [42]	Radiologic	Random Forest	Internal		0.64	67.0	70.0	58.0
<i>DSS/CSS</i>								
Csúry 2024 [48]	Pathologic, Clinical	Naïve Bayes	Independent		0.85		77.0	
Peng 2022 [22]	Clinical	Random Forest	Independent	0.84				
Lu 2017 [49]	Pathologic	Quadratic Discriminant Analysis	Internal		0.87	88.0	78.0	93.0
Wang 2023 [50]	Clinical	Random Survival Forest	Internal		0.80			
Adeoye 2022 [29]	Clinical, Pathologic	DeepSurv	Internal	0.89				
Tseng 2015 [45]	Clinical, Pathologic	Decision tree	Internal			81.7		

4. Discussion

This systematic review evaluated machine learning models developed for prognosis prediction in oral cavity squamous cell carcinoma and found that ML-based approaches generally demonstrated favorable discriminatory ability across multiple clinically relevant outcomes, including overall survival, recurrence, disease-free survival, and disease-specific survival. Independently validated models, particularly those derived from larger multicenter or registry-based cohorts with lower overall risk of bias, most often reported AUCs in the approximate range of 0.70–0.90, supporting the potential utility of ML-assisted prognostic stratification in OCSCC. AUC was emphasized in this review because it was the most consistently reported performance metric across studies and therefore allowed the broadest comparison of model discrimination. However, AUC alone does not assess calibration, clinical utility, or the impact of model-guided decision making, and should not be interpreted as a comprehensive measure of model quality. Findings remained robust when examined qualitatively according to validation strategy and PROBAST + AI risk-of-bias classification, with extreme near-perfect AUC values largely confined to small, internally validated, higher-risk-of-bias studies suggestive of optimism bias.

Across the literature, a wide range of machine learning architectures were utilized, including ensemble learning methods, random forests, SVMs, gradient boosting approaches, neural networks, and deep learning-based survival frameworks. Most models incorporated multimodal combinations of clinical, pathologic, radiologic, immunologic, or genomic variables, reflecting the increasing emphasis on integrated prognostic modeling in oncology. Notably, while multimodal models frequently demonstrated the highest performance, many relied on radiomic, genomic, immunologic, or digital pathology features that lack standardized acquisition and analytic frameworks. This may limit reproducibility and external generalizability.

These findings are broadly consistent with prior systematic reviews and meta-analyses examining artificial intelligence and prediction models in head and neck cancer, which have similarly reported strong apparent discrimination alongside substantial methodological heterogeneity [8,51]. Adeoye et al. identified promising performance of machine learning models for oral cavity cancer outcomes but highlighted limitations related to heterogeneous methodologies, small datasets, and limited validation [52]. Similar conclusions were reached by subsequent reviews, which emphasized the growing complexity of machine learning approaches alongside persistent concerns regarding reproducibility, external validation, and clinical implementation [8,10,51,53]. Our findings largely corroborate these observations. However, compared with earlier reviews, the current study includes a substantially larger and more contemporary body of evidence, captures recent advances in radiomic, genomic, immunologic, and multimodal prediction models, and applies the PROBAST + AI framework to systematically assess methodological quality and risk of bias. Furthermore, by differentiating independently validated models from internally validated models, our review provides additional insight into model generalizability and identifies a subset of lower-risk-of-bias models with greater potential for clinical translation.

An important observation was that models trained and evaluated using larger multicenter cohorts with independent validation tended to demonstrate more modest but clinically plausible performance estimates, while models developed in small single-institution datasets occasionally reported near-perfect AUCs. This pattern mirrors concerns raised in broader oncology ML research, where overly optimistic internal validation results may reflect overfitting, data leakage, or inadequate separation of training and test sets rather than true generalizability. Accordingly, future efforts should prioritize rigorous external validation of existing models rather than continued development of internally validated algorithms.

Additionally, outcome definitions, predictor selection, and validation methodology were also heterogeneous across studies, limiting direct comparability of results. Some models focused exclusively on conventional clinicopathologic variables, while others incorporated radiomics, digital pathology, immunologic signatures, or genomic data. Although multimodal models frequently achieved high discrimination, their reliance on non-routine imaging or molecular pipelines, variable feature extraction workflows, and complex preprocessing may constrain reproducibility and implementation outside specialized centers. In addition, relatively few studies reported calibration, decision-curve analysis, or clinical utility metrics, which are essential for understanding the practical impact of prognostic models beyond discrimination alone.

From a clinical standpoint, the actionable implications of the current literature remain limited. Although several ML models demonstrated acceptable discrimination for survival, recurrence, and nodal metastasis prediction, few studies evaluated whether model outputs would meaningfully alter treatment selection, adjuvant therapy escalation or de-escalation, surveillance intensity, or patient counseling beyond existing clinicopathologic risk stratification. In addition, relatively few models were directly compared with standard-of-care prognostic systems such as AJCC TNM staging, established nomograms, or conventional Cox-based models. Where comparisons were performed, ML models generally showed comparable rather than clearly superior performance. Therefore, the available evidence does not yet demonstrate consistent incremental clinical utility of ML over established prognostic frameworks. Future studies should explicitly evaluate whether ML-assisted risk stratification improves decision-making, calibration, net benefit, and downstream patient outcomes when compared with current standard prognostic systems.

An additional and important observation from this review was the near-complete absence of calibration assessment and decision-curve analysis among included studies. Although discrimination metrics such as AUC and C-index were frequently reported, these measures alone do not establish whether predicted risks are well calibrated or whether model-guided decisions would improve clinical outcomes. Calibration and decision-curve analyses provide complementary information regarding model reliability and clinical utility and are increasingly recognized as essential components of prediction model evaluation. Their limited use represents a significant methodological and reporting gap within the current OCSCC machine learning literature. Future studies should routinely incorporate and report these metrics to improve transparency, facilitate model comparison, and support eventual clinical implementation.

This review has several limitations. First, substantial heterogeneity in study design, predictor selection, outcome definitions, and reporting of performance metrics precluded formal quantitative meta-analysis of AUCs or C-indices and necessitated a primarily descriptive synthesis of prognostic performance. Additionally, extracting the highest-performing model from each study may introduce optimism bias and overestimate the reported predictive performance across the literature. Although studies were stratified according to the clinical outcome being predicted, further subgroup analyses based on model architecture, predictor modality, or validation approach were not feasible because of substantial overlap between categories, inconsistent reporting, and the limited number of studies within many potential subgroups. In addition, while AUC quantifies a model's ability to distinguish between patients who do and do not experience an outcome, it does not assess calibration, clinical utility, or whether risk predictions would meaningfully influence management decisions. Consequently, high AUC values should not be interpreted as evidence of clinical usefulness or superiority over existing prognostic tools. Second, incomplete reporting of variance estimates, missing data handling, and potential cohort overlap limited assessment of methodological rigor in some studies and may have biased risk-of-bias judgments. Third, most included studies were retrospective and developed models in single-institution or regional datasets, which may limit generalizability to other practice settings. Fourth, publication bias is possible, as studies reporting strong model performance may be more likely to be published, while negative or inconclusive ML models may remain underreported. Finally, although we attempted prospective registration through PROSPERO, no registration number was issued because PROSPERO does not currently accept systematic reviews focused on diagnostic test accuracy and related prediction model methodologies. This should be considered when interpreting the findings.

Despite these limitations, the present review highlights several priorities for future work. Prospective, multicenter studies with prespecified analysis plans and independent external validation are needed to generate more reliable estimates of ML model performance for OCSCC prognosis. Adoption of standardized reporting frameworks tailored to prediction models using artificial intelligence, such as TRIPOD-AI and PROBAST + AI, will be critical to improve transparency, facilitate reproducibility, and allow robust comparative appraisal of competing models. Future studies should routinely report both discrimination and calibration, incorporate decision-analytic measures of clinical utility, and, where feasible, provide open-source code, model parameters, and harmonized feature definitions to enable independent validation and updating. Ultimately, integration of rigorously validated ML models into clinical decision support will require not only technical performance but also consideration of workflow integration, interpretability, and real-world impact on treatment decisions and patient outcomes.

5. Conclusions

The present review provides a focused synthesis of machine learning prognostic modeling in OCSCC and demonstrates that contemporary ML models, particularly multimodal approaches integrating clinical, pathologic, radiologic, and molecular data, generally show favorable prognostic performance across survival and recurrence outcomes. However, substantial heterogeneity in methodology, reporting quality, and validation practices continues to limit generalizability and clinical implementation. Future studies should prioritize prospective multicenter validation, transparent reporting standards, calibration assessment, and reproducible modeling pipelines.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/cancers18142261/s1>, Figure S1: Definitions of these performance metrics; File S1: Guideline for Reporting Systematic Review; Table S1: Search Terms; Table S2: Characteristics of Included Studies; Table S3: Risk of Bias Assessment of Model Development with PROBAST + AI Development Tool; Table S4: Risk of Bias Assessment of Model Validation with PROBAST + AI Evaluation Tool.

Author Contributions: Conceptualization, S.Y.G., J.M.H., S.A.N. and J.G.N.; methodology, S.Y.G., J.M.H., S.A.N. and J.G.N.; software, S.A.N.; validation, S.A.N. and J.G.N.; formal analysis, S.Y.G., J.M.H., B.S.M. and S.A.N.; resources, S.A.N. and J.G.N.; data curation, S.Y.G., J.M.H., B.S.M. and S.A.N.; writing—original draft preparation, S.Y.G., J.M.H. and B.S.M.; writing—review and editing, S.Y.G., J.M.H., B.S.M., S.A.N. and J.G.N.; visualization, S.Y.G. and J.M.H.; supervision, S.A.N. and J.G.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable. This study is a systematic review of published literature and did not involve direct interaction with human participants or access to identifiable private data.

Informed Consent Statement: Not applicable. Patient consent was waived because this study is a systematic review of previously published literature and did not involve the direct recruitment of participants, interaction with patients, or collection of identifiable individual-level data.

Data Availability Statement: No new datasets were generated during this study. The data supporting the findings of this systematic review are available from the corresponding author upon reasonable request and can also be obtained from the original studies cited in the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
AJCC	American Joint Committee on Cancer
AUC	Area Under the Receiver Operating Characteristic Curve
BNI	Bone Invasion
C-index	Concordance Index
CI	Confidence Interval
CSS	Cancer-Specific Survival
DFS	Disease-Free Survival
DSS	Disease-Specific Survival
ECE	Extranodal Extension
ML	Machine Learning
OCSCC	Oral Cavity Squamous Cell Carcinoma
OS	Overall Survival

PNI	Perineural Invasion
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROBAST + AI	Prediction Model Risk of Bias Assessment Tool for Artificial Intelligence
RFS	Recurrence-Free Survival
ROC	Receiver Operating Characteristic
SVM	Support Vector Machine
TNM	Tumor, Node, Metastasis
LVI	Lymphovascular Invasion
CT	Computed Tomography
MRI	Magnetic Resonance Imaging
PET	Positron Emission Tomography
FDG-PET	Fluorodeoxyglucose Positron Emission Tomography
XGBoost	Extreme Gradient Boosting
MeSH	Medical Subject Headings
TRIPOD-AI	Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis–Artificial Intelligence Extension

References

- Coletta, R.D.; Yeudall, W.A.; Salo, T. Current Trends on Prevalence, Risk Factors and Prevention of Oral Cancer. *Front. Oral Health* **2024**, *5*, 1505833. <https://doi.org/10.3389/froh.2024.1505833>.
- Wang, Z.; Li, Y.; Feng, J.; Wu, L.; Li, T.; Zhu, X.; Liang, L. Survival and Pathological Outcomes of Neoadjuvant Chemoimmunotherapy Followed by Surgery in Locally Advanced Oral Squamous Cell Carcinoma. *Cancer Control* **2026**, *33*, 10732748261453684. <https://doi.org/10.1177/10732748261453684>.
- Imbesi Bellantoni, M.; Picciolo, G.; Pirrotta, I.; Irrera, N.; Vaccaro, M.; Vaccaro, F.; Squadrito, F.; Pallio, G. Oral Cavity Squamous Cell Carcinoma: An Update of the Pharmacological Treatment. *Biomedicines* **2023**, *11*, 1112. <https://doi.org/10.3390/biomedicines11041112>.
- Montero, P.H.; Yu, C.; Palmer, F.L.; Patel, P.D.; Ganly, I.; Shah, J.P.; Shaha, A.R.; Boyle, J.O.; Kraus, D.H.; Singh, B.; et al. Nomograms for Preoperative Prediction of Prognosis in Patients with Oral Cavity Squamous Cell Carcinoma. *Cancer* **2014**, *120*, 214–221. <https://doi.org/10.1002/cncr.28407>.
- Zhou, J.; Li, H.; Cheng, B.; Cao, R.; Zou, F.; Yang, D.; Liu, X.; Song, M.; Wu, T. Derivation and Validation of a Prognostic Scoring Model Based on Clinical and Pathological Features for Risk Stratification in Oral Squamous Cell Carcinoma Patients: A Retrospective Multicenter Study. *Front. Oncol.* **2021**, *11*, 652553. <https://doi.org/10.3389/fonc.2021.652553>.
- Karadaghy, O.A.; Shew, M.; New, J.; Bur, A.M. Development and Assessment of a Machine Learning Model to Help Predict Survival Among Patients with Oral Squamous Cell Carcinoma. *JAMA Otolaryngol. Head Neck Surg.* **2019**, *145*, 1115–1120. <https://doi.org/10.1001/jamaoto.2019.0981>.
- Vollmer, A.; Hartmann, S.; Vollmer, M.; Shavlokhova, V.; Brands, R.C.; Kübler, A.; Wollborn, J.; Hassel, F.; Couillard-Despres, S.; Lang, G.; et al. Multimodal Artificial Intelligence-Based Pathogenomics Improves Survival Prediction in Oral Squamous Cell Carcinoma. *Sci. Rep.* **2024**, *14*, 5687. <https://doi.org/10.1038/s41598-024-56172-5>.
- Moharrami, M.; Azimian Zavareh, P.; Watson, E.; Singhal, S.; Johnson, A.E.W.; Hosni, A.; Quinonez, C.; Glogauer, M. Prognosing Post-Treatment Outcomes of Head and Neck Cancer Using Structured Data and Machine Learning: A Systematic Review. *PLoS ONE* **2024**, *19*, e0307531. <https://doi.org/10.1371/journal.pone.0307531>.
- Rao, K.N.; Fernandez-Alvarez, V.; Guntinas-Lichius, O.; Sreeram, M.P.; de Bree, R.; Kowalski, L.P.; Forastiere, A.; Pace-Asciak, P.; Rodrigo, J.P.; Saba, N.F.; et al. The Limitations of Artificial Intelligence in Head and Neck Oncology. *Adv. Ther.* **2025**, *42*, 2559–2568. <https://doi.org/10.1007/s12325-025-03198-4>.
- Mäkitie, A.A.; Alabi, R.O.; Ng, S.P.; Takes, R.P.; Robbins, K.T.; Ronen, O.; Shaha, A.R.; Bradley, P.J.; Saba, N.F.; Nuyts, S.; et al. Artificial Intelligence in Head and Neck Cancer: A Systematic Review of Systematic Reviews. *Adv. Ther.* **2023**, *40*, 3360–3380. <https://doi.org/10.1007/s12325-023-02527-9>.
- The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews|The BMJ. Available online: <https://www.bmj.com/content/372/bmj.n71> (accessed on 20 May 2026).
- Dhiman, P.; Ma, J.; Andaur Navarro, C.L.; Speich, B.; Bullock, G.; Damen, J.A.A.; Hooft, L.; Kirtley, S.; Riley, R.D.; Van Calster, B.; et al. Methodological Conduct of Prognostic Prediction Models Developed Using Machine Learning in Oncology: A Systematic Review. *BMC Med. Res. Methodol.* **2022**, *22*, 101. <https://doi.org/10.1186/s12874-022-01577-x>.

13. Moons, K.G.M.; Damen, J.A.A.; Kaul, T.; Hooft, L.; Navarro, C.A.; Dhiman, P.; Beam, A.L.; Calster, B.V.; Celi, L.A.; Denaxas, S.; et al. PROBAST+AI: An Updated Quality, Risk of Bias, and Applicability Assessment Tool for Prediction Models Using Regression or Artificial Intelligence Methods. *BMJ* **2025**, *388*, e082505. <https://doi.org/10.1136/bmj-2024-082505>.
14. Somyanonthanakul, R.; Warin, K.; Chaowchuen, S.; Jinaporntham, S.; Panichkitkosolkul, W.; Suebnukarn, S. Survival Estimation of Oral Cancer Using Fuzzy Deep Learning. *BMC Oral. Health* **2024**, *24*, 519. <https://doi.org/10.1186/s12903-024-04279-6>.
15. Lee, J.; Lin, J.-B.; Lin, W.-C.; Jan, Y.-T.; Leu, Y.-S.; Chen, Y.-J.; Wu, K.-P. Identifying Threshold of CT-Defined Muscle Loss after Radiotherapy for Survival in Oral Cavity Cancer Using Machine Learning. *Eur. Radiol.* **2025**, *35*, 4289–4299. <https://doi.org/10.1007/s00330-024-11303-4>.
16. Alabi, R.O.; Elmusrati, M.; Leivo, I.; Almangush, A.; Mäkitie, A.A. Collaborative Machine Learning-Guided Overall Survival Prediction of Oral Squamous Cell Carcinoma. *Acta Otolaryngol.* **2026**, *146*, 133–140. <https://doi.org/10.1080/00016489.2024.2437012>.
17. Mes, S.W.; van Velden, F.H.P.; Peltenburg, B.; Peeters, C.F.W.; Te Beest, D.E.; van de Wiel, M.A.; Mekke, J.; Mulder, D.C.; Martens, R.M.; Castelijns, J.A.; et al. Outcome Prediction of Head and Neck Squamous Cell Carcinoma by MRI Radiomic Signatures. *Eur. Radiol.* **2020**, *30*, 6311–6321. <https://doi.org/10.1007/s00330-020-06962-y>.
18. Han, X.; Sun, T.; Dai, Y.; Yun, R.; Wang, H.; Jia, J.; Feng, X.; Jiao, M.; Hou, M.; Yue, M.; et al. Development of a Cancer Metastasis-Associated Risk Model via Multi-Machine-Learning Algorithms for Prognostic Risk Evaluation and Clinical Application in Oral Squamous Cell Carcinoma. *J. Transl. Med.* **2025**, *23*, 1344. <https://doi.org/10.1186/s12967-025-07336-y>.
19. Deepali; Goel, N.; Khandnor, P. DeepOmicsSurv: A Deep Learning-Based Model for Survival Prediction of Oral Cancer. *Discov. Oncol.* **2025**, *16*, 614. <https://doi.org/10.1007/s12672-025-02346-0>.
20. Diao, P.; Jiang, Y.; Li, Y.; Wu, X.; Li, J.; Zhou, C.; Jiang, L.; Zhang, W.; Yan, E.; Zhang, P.; et al. Immune Landscape and Subtypes in Primary Resectable Oral Squamous Cell Carcinoma: Prognostic Significance and Predictive of Therapeutic Response. *J. Immunother. Cancer* **2021**, *9*, e002434. <https://doi.org/10.1136/jitc-2021-002434>.
21. Kim, D.W.; Lee, S.; Kwon, S.; Nam, W.; Cha, I.-H.; Kim, H.J. Deep Learning-Based Survival Prediction of Oral Cancer Patients. *Sci. Rep.* **2019**, *9*, 6994. <https://doi.org/10.1038/s41598-019-43372-7>.
22. Peng, J.; Lu, Y.; Chen, L.; Qiu, K.; Chen, F.; Liu, J.; Xu, W.; Zhang, W.; Zhao, Y.; Yu, Z.; et al. The Prognostic Value of Machine Learning Techniques versus Cox Regression Model for Head and Neck Cancer. *Methods* **2022**, *205*, 123–132. <https://doi.org/10.1016/j.ymeth.2022.07.001>.
23. Alabi, R.O.; Elmusrati, M.; Leivo, I.; Almangush, A.; Mäkitie, A.A. Advanced-Stage Tongue Squamous Cell Carcinoma: A Machine Learning Model for Risk Stratification and Treatment Planning. *Acta Oto-Laryngol.* **2023**, *143*, 206–214. <https://doi.org/10.1080/00016489.2023.2172208>.
24. Alabi, R.O.; Mäkitie, A.A.; Pirinen, M.; Elmusrati, M.; Leivo, I.; Almangush, A. Comparison of Nomogram with Machine Learning Techniques for Prediction of Overall Survival in Patients with Tongue Cancer. *Int. J. Med. Inform.* **2021**, *145*, 104313. <https://doi.org/10.1016/j.ijmedinf.2020.104313>.
25. Li, L.; Pu, C.; Jin, N.; Zhu, L.; Hu, Y.; Cascone, P.; Tao, Y.; Zhang, H. Prediction of 5-Year Overall Survival of Tongue Cancer Based Machine Learning. *BMC Oral. Health* **2023**, *23*, 567. <https://doi.org/10.1186/s12903-023-03255-w>.
26. Huang, S.-N.; Li, G.-S.; Zhou, X.-G.; Chen, X.-Y.; Yao, Y.-X.; Zhang, X.-G.; Liang, Y.; Li, M.-X.; Chen, G.; Huang, Z.-G.; et al. Identification of an Immune Score-Based Gene Panel with Prognostic Power for Oral Squamous Cell Carcinoma. *Med. Sci. Monit.* **2020**, *26*, e922854. <https://doi.org/10.12659/MSM.922854>.
27. Lin, W.; Shu, D.; Geng, N.; Zhang, D.; Feng, C.; Chen, S. DeepSurv-Based Deep Learning Model for Survival Prediction and Personalized Treatment Recommendation in Tongue Squamous Cell Carcinoma. *J. Cranio-Maxillofac. Surg.* **2025**, *53*, 1334–1343. <https://doi.org/10.1016/j.jcms.2025.05.005>.
28. Tan, J.Y.; Adeoye, J.; Thomson, P.; Sharma, D.; Ramamurthy, P.; Choi, S.-W. Predicting Overall Survival Using Machine Learning Algorithms in Oral Cavity Squamous Cell Carcinoma. *Anticancer Res.* **2022**, *42*, 5859–5866. <https://doi.org/10.21873/anticancer.16094>.
29. Adeoye, J.; Hui, L.; Koohi-Moghadam, M.; Tan, J.Y.; Choi, S.-W.; Thomson, P. Comparison of Time-to-Event Machine Learning Models in Predicting Oral Cavity Cancer Prognosis. *Int. J. Med. Inform.* **2022**, *157*, 104635. <https://doi.org/10.1016/j.ijmedinf.2021.104635>.
30. Nezamabadi Farahani, L.; Kazemnejad, A.; Afrasiabi, M.; Tapak, L. Unlocking The Potential of Hybrid Models for Prognostic Biomarker Discovery in Oral Cancer Survival Analysis: A Retrospective Cohort Study. *Cell J.* **2025**, *26*, 688–699. <https://doi.org/10.22074/cellj.2025.2034704.1618>.

31. Rosado, P.; Lequerica-Fernández, P.; Villallaín, L.; Peña, I.; Sanchez-Lasheras, F.; de Vicente, J.C. Survival Model in Oral Squamous Cell Carcinoma Based on Clinicopathological Parameters, Molecular Markers and Support Vector Machines. *Expert. Syst. Appl.* **2013**, *40*, 4770–4776. <https://doi.org/10.1016/j.eswa.2013.02.032>.
32. Adachi, M.; Taki, T.; Kojima, M.; Sakamoto, N.; Matsuura, K.; Hayashi, R.; Tabuchi, K.; Ishikawa, S.; Ishii, G.; Sakashita, S. Predicting Lymph Node Recurrence in cT1-2N0 Tongue Squamous Cell Carcinoma: Collaboration between Artificial Intelligence and Pathologists. *J. Pathol. Clin. Res.* **2024**, *10*, e12392. <https://doi.org/10.1002/2056-4538.12392>.
33. Cai, Y.; Xie, Y.; Zhang, S.; Wang, Y.; Chen, J.; Huang, Z. Prediction of Postoperative Recurrence of Oral Cancer by Artificial Intelligence Model: Multilayer Perceptron. *Head. Neck* **2023**, *45*, 3053–3066. <https://doi.org/10.1002/hed.27533>.
34. Chen, Z.; Chen, L.; Li, S.; Xia, P.; Lam, A.K.; Qiao, J.; Liu, Y.; Qiao, B. Integrated Analysis of Oral Rinse-Derived and Plasma Circulating Tumour DNA for Mutation Profiling and Outcome Prediction with Oral Squamous Cell Carcinoma. *NPJ Precis. Oncol.* **2025**, *9*, 183. <https://doi.org/10.1038/s41698-025-00976-9>.
35. Vidiri, A.; Dolcetti, V.; Mazzola, F.; Lucchese, S.; Laganaro, F.; Piludu, F.; Pellini, R.; Covello, R.; Marzi, S. MRI in Oral Tongue Squamous Cell Carcinoma: A Radiomic Approach in the Local Recurrence Evaluation. *Curr. Oncol.* **2025**, *32*, 116. <https://doi.org/10.3390/curroncol32020116>.
36. Fatapour, Y.; Abiri, A.; Kuan, E.C.; Brody, J.P. Development of a Machine Learning Model to Predict Recurrence of Oral Tongue Squamous Cell Carcinoma. *Cancers* **2023**, *15*, 2769. <https://doi.org/10.3390/cancers15102769>.
37. Huang, S.Y.; Hsu, R.J.; Liu, D.W.; Hsu, W.L. Using a Machine Learning Algorithm and Clinical Data to Predict the Risk Factors of Disease Recurrence after Adjuvant Treatment of Advanced-Stage Oral Cavity Cancer. *Tzu Chi Med. J.* **2025**, *37*, 91–98. https://doi.org/10.4103/tcmj.tcmj_56_24.
38. Alabi, R.O.; Elmusrati, M.; Sawazaki-Calone, I.; Kowalski, L.P.; Haglund, C.; Coletta, R.D.; Mäkitie, A.A.; Salo, T.; Almangush, A.; Leivo, I. Comparison of Supervised Machine Learning Classification Techniques in Prediction of Locoregional Recurrences in Early Oral Tongue Cancer. *Int. J. Med. Inform.* **2020**, *136*, 104068. <https://doi.org/10.1016/j.ijmedinf.2019.104068>.
39. Bourdillon, A.T.; Shah, H.P.; Cohen, O.; Hajek, M.A.; Mehra, S. Novel Machine Learning Model to Predict Interval of Oral Cancer Recurrence for Surveillance Stratification. *Laryngoscope* **2023**, *133*, 1652–1659. <https://doi.org/10.1002/lary.30351>.
40. Alabi, R.O.; Elmusrati, M.; Sawazaki-Calone, I.; Kowalski, L.P.; Haglund, C.; Coletta, R.D.; Mäkitie, A.A.; Salo, T.; Leivo, I.; Almangush, A. Machine Learning Application for Prediction of Locoregional Recurrences in Early Oral Tongue Cancer: A Web-Based Prognostic Tool. *Virchows Arch.* **2019**, *475*, 489–497. <https://doi.org/10.1007/s00428-019-02642-5>.
41. Salehi, A.M.; Wang, L.; Gu, X.; Coates, P.J.; Norberg Spaak, L.; Sgaramella, N.; Nylander, K. Patients with Oral Tongue Squamous Cell Carcinoma and Co-existing Diabetes Exhibit Lower Recurrence Rates and Improved Survival: Implications for Treatment. *Oncol. Lett.* **2024**, *27*, 142. <https://doi.org/10.3892/ol.2024.14275>.
42. Pfahler, E.; Schindele, A.; Dierks, A.; Busse, C.; Brumberg, J.; Kübler, A.C.; Buck, A.K.; Linz, C.; Lapa, C.; Brands, R.C.; et al. Value of PET Radiomic Features for Diagnosis and Recurrence Prediction of Newly Diagnosed Oral Squamous Cell Carcinoma. *Sci. Rep.* **2025**, *15*, 17475. <https://doi.org/10.1038/s41598-025-02305-3>.
43. Alkhadar, H.; Macluskey, M.; White, S.; Ellis, I.; Gardner, A. Comparison of Machine Learning Algorithms for the Prediction of Five-Year Survival in Oral Squamous Cell Carcinoma. *J. Oral. Pathol. Med.* **2021**, *50*, 378–384. <https://doi.org/10.1111/jop.13135>.
44. Fujima, N.; Andreu-Arasa, V.C.; Meibom, S.K.; Mercier, G.A.; Salama, A.R.; Truong, M.T.; Sakai, O. Prediction of the Treatment Outcome Using Machine Learning with FDG-PET Image-Based Multiparametric Approach in Patients with Oral Cavity Squamous Cell Carcinoma. *Clin. Radiol.* **2021**, *76*, e711.e1–711.e7. <https://doi.org/10.1016/j.crad.2021.03.017>.
45. Tseng, W.T.; Chiang, W.F.; Liu, S.Y.; Roan, J.; Lin, C.N. The Application of Data Mining Techniques to Oral Cancer Prognosis. *J. Med. Syst.* **2015**, *39*, 59. <https://doi.org/10.1007/s10916-015-0241-3>.
46. Mei, X.; Luo, W.; Duan, W.; Guo, Z.; Lao, X.; Zhang, S.; Yang, L.; Zeng, B.; Gong, J.; Deng, W.; et al. Development and Validation of Machine Learning Models for Predicting Tumor Progression in OSCC. *Oral. Dis.* **2025**, *31*, 426–434. <https://doi.org/10.1111/odi.15159>.
47. Chu, C.S.; Lee, N.P.; Adeoye, J.; Thomson, P.; Choi, S.W. Machine Learning and Treatment Outcome Prediction for Oral Cancer. *J. Oral. Pathol. Med.* **2020**, *49*, 977–985. <https://doi.org/10.1111/jop.13089>.
48. Csúry, T.D.; Csúry, A.Z.; Balk, M.; Kist, A.M.; Rupp, R.; Mueller, S.K.; Sievert, M.; Iro, H.; Eckstein, M.; Gostian, A.O. The Modified Polsby-Popper Score, a Novel Quantitative Histomorphological Biomarker and Its Potential to Predict Lymph Node Positivity and Cancer-Specific Survival in Oral Tongue Squamous Cell Carcinoma. *Cancer Med.* **2024**, *13*, e6824. <https://doi.org/10.1002/cam4.6824>.

49. Lu, C.; Lewis, J.S.J.r.; Dupont, W.D.; Plummer, W.D.J.r.; Janowczyk, A.; Madabhushi, A. An Oral Cavity Squamous Cell Carcinoma Quantitative Histomorphometric-Based Image Classifier of Nuclear Morphology Can Risk Stratify Patients for Disease-Specific Survival. *Mod. Pathol.* **2017**, *30*, 1655–1665. <https://doi.org/10.1038/modpathol.2017.98>.
50. Wang, N.; Lin, Y.; Song, H.; Huang, W.; Huang, J.; Shen, L.; Chen, F.; Liu, F.; Wang, J.; Qiu, Y.; et al. Development and Validation of a Model for the Prediction of Disease-Specific Survival in Patients with Oral Squamous Cell Carcinoma: Based on Random Survival Forest Analysis. *Eur. Arch. Otorhinolaryngol.* **2023**, *280*, 5049–5057. <https://doi.org/10.1007/s00405-023-08087-6>.
51. Wang, T.-W.; Wang, C.-K.; Hong, J.-S.; Lin, Y.-H.; Wang, S.-Y.; Lu, C.-F.; Wu, Y.-T. Prognostic Power of Radiomics in Head and Neck Cancers: Insights from a Meta-Analysis. *Comput. Methods Programs Biomed.* **2025**, *262*, 108683. <https://doi.org/10.1016/j.cmpb.2025.108683>.
52. Adeoye, J.; Tan, J.Y.; Choi, S.-W.; Thomson, P. Prediction Models Applying Machine Learning to Oral Cavity Cancer Outcomes: A Systematic Review. *Int. J. Med. Inform.* **2021**, *154*, 104557. <https://doi.org/10.1016/j.ijmedinf.2021.104557>.
53. Warin, K.; Suebnukarn, S. Deep Learning in Oral Cancer—A Systematic Review. *BMC Oral. Health* **2024**, *24*, 212. <https://doi.org/10.1186/s12903-024-03993-5>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.