ORL

Systematic Review

Machine Learning Algorithms as a Computer-Assisted Decision Tool for Oral Cancer Prognosis and Management Decisions: A Systematic Review

Carlos M. Chiesa-Estomba[a, b]   Manuel Graña[c]   Alfonso Medela[d]
Jon A. Sistiaga-Suarez[a]   Jerome R. Lechien[b, e]   Christian Calvo-Henriquez[b, f]
Miguel Mayo-Yanez[b, g]   Luigi Angelo Vaira[h]   Alberto Grammatica[i]
Giovanni Cammaroto[b, j]   Tareck Ayad[b, k]   Johannes J. Fagan[l]

[a]Otorhinolaryngology – Head & Neck Surgery Department, Hospital Universitario Donostia, Biodonostia Health Research Institute, San Sebastian, Spain; [b]Head & Neck Study Group of Young-Otolaryngologists of the International Federations of Oto-Rhino-Laryngological Societies (YO-IFOS), San Sebastian, Spain; [c]Computational Intelligence Group, Facultad de Informatica UPV/EHU, San Sebastian, Spain; [d]LEGIT Health, Bilbao, Spain; [e]Department of Human Anatomy & Experimental Oncology, University of Mons, Mons, Belgium; [f]Department of Otolaryngology – Hospital Complex of Santiago de Compostela, Santiago de Compostela, Spain; [g]Otorhinolaryngology – Head and Neck Surgery Department, Complexo Hospitalario Universitario A Coruña (CHUAC), A Coruña, Spain; [h]Maxillofacial Surgery Unit, University Hospital of Sassari, Sassari, Italy; [i]Department of Otorhinolaryngology – Head and Neck Surgery, University of Brescia, Brescia, Italy; [j]Department of Otolaryngology–Head & Neck Surgery, Morgagni Pierantoni Hospital, Forli, Italy; [k]Division of Otolaryngology–Head & Neck Surgery, Centre Hospitalier de l'Université de Montréal, Montreal, QC, Canada; [l]Division of Otolaryngology, Groote Schuur Hospital, University of Cape Town, Cape Town, South Africa

Abstract

Introduction: Despite multiple prognostic indicators described for oral cavity squamous cell carcinoma (OCSCC), its management still continues to be a matter of debate. Machine learning is a subset of artificial intelligence that enables computers to learn from historical data, gather insights, and make predictions about new data using the model learned. Therefore, it can be a potential tool in the field of head and neck cancer. **Methods:** We conducted a systematic review. **Results:** A total of 81 manuscripts were revised, and 46 studies met the inclusion criteria. Of these, 38 were excluded for the following reasons: use of a classical statistical method ($N = 16$), nonspecific for OCSCC ($N = 15$), and not being related to OCSCC survival ($N = 7$). In total, 8 studies were included in the final analysis. **Conclusions:** ML has the potential to significantly advance research in the field of OCSCC. Advantages are related to the use and training of ML models because of their capability to continue training continuously when more data become available. Future ML research will allow us to improve and democratize the application of algorithms to improve the prediction of cancer prognosis and its management worldwide.

© 2022 S. Karger AG, Basel

Correspondence to:
Carlos M. Chiesa-Estomba, chiesaestomba86@gmail.com

Introduction

Oral cavity squamous cell carcinoma (OCSCC) has experienced a gradual increase all over the world in the past few decades, and currently, is the most common type of head and neck malignant tumor and the 8th leading cause of death worldwide with 300,000 new cases and 145,000 deaths per year worldwide [1, 2]. Surgery continues to be the first-line therapy, and the National Comprehensive Cancer Network recommends primary surgical management in both early- and late-stage disease [3], i.e., excision of the primary tumor with or without neck dissection [4, 5].

Despite the multiple prognostic indicators described for OCSCC, lymph node metastasis (LNM) continues to be the most relevant factor [5–10]. Hence, neck assessment is considered mandatory for risk stratification due to the high risk of occult LNM (OLNM). Consequently, different strategies have been described to manage the risk of OLNM, such as close clinical follow-up, reserving therapeutic neck dissection just for those patients who subsequently develop an LNM, sentinel lymph node biopsy, or elective neck dissection. The selection of patients that will benefit from elective neck dissection has been a matter of debate over decades, and one has to balance the rate of pathologically confirmed N0 patients who undergo unnecessary surgery and the long-term sequelae associated with neck dissection, against the benefits of surgically addressing occult nodal metastases [11].

More recently, depth of invasion (DOI) of the primary oral cancer has become the most commonly used histopathologic variable to predict the risk of occult nodal metastasis [12]. Other prognostic factors described in the literature include tumor location, histologic grade, sex, age, extracapsular spread, perineural invasion, lymphovascular invasion, recurrence, or tobacco consumption [13–18]. These factors have been combined to create multivariate regression models and nomograms to predict survival [19–21]. However, these factors have not been widely adopted because of their low predictive accuracy and difficulty to use in daily practice.

Machine learning (ML) is a subset of artificial intelligence and with the increasing availability of large national databases and computing power, the amount of potential input data has increased, and it has become necessary to explore novel approaches of data analysis, to achieve more accurate and precise predictions [22–29]. In this regard, ML represents an alternative way of developing cancer survival prediction models.

The main objective of using ML techniques in medicine is to produce a model that can be used to predict the medical outcome (either diagnosis or the prognosis) of the patient from multivariate data. All these algorithms were designed to learn and find statistical regularities or common patterns from each dataset. The aim of this manuscript was therefore to review and provide a comprehensive summary of all the evidence related to the use of ML algorithms as a noninvasive tool to evaluate the OCSCC prognosis and stratify the risk of recurrence.

Methods

The systematic approach employed for the search strategy in peer-reviewed journals on the use of ML algorithms to evaluate prognosis in OCSCC patients was based on the recommendations of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement, reflected in the PRISMA checklist file included [30]. Inclusion criteria were based on the population, intervention, comparison, outcome, timing, and setting framework [31]. The heterogeneity among studies, mainly due to the absence of randomization, limited our ability to statistically combine data into a formal meta-analysis.

Eligibility Criteria

All prospective, retrospective, controlled, or uncontrolled studies published in peer-reviewed English language journals that investigated the role of ML in OCSCC patients and described an algorithm to evaluate prognosis were considered.

Participant's Inclusion/Exclusion Criteria

Studies were considered for analysis that reported results of patients aged >18 years and described the use of ML algorithms to evaluate prognosis in OCSCC cases. Studies including other head and neck subsites or in patients aged <18 years or not related to ML as a tool to investigate prognosis were excluded.

Intervention and Comparison

This study investigated the role of ML in determining prognosis in OCSCC cases by evaluating predictors such as demographic data, clinical symptoms, imaging, features, pathological data, or genomic data. ML algorithms were compared with traditional statistical methods as well as other clinical prognostic instruments.

Outcomes

The primary outcome evaluated was progression or recurrence rates, through area under the curve (AUC), and algorithm accuracy analysis.

Timing

In studies evaluating overall survival, a minimum median follow-up time of 3 years after treatment was used to evaluate prognosis.

Setting

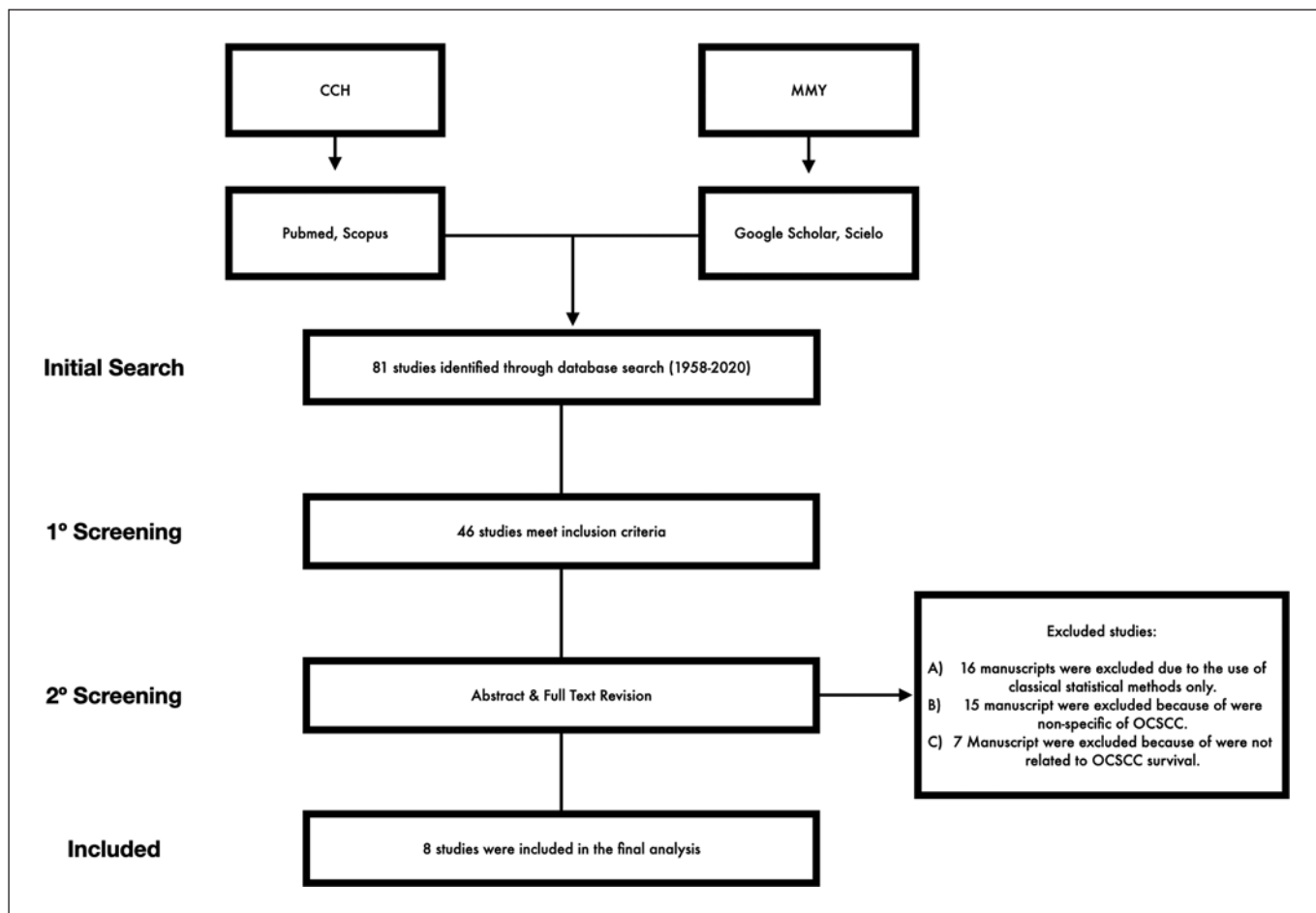Tertiary academic and nonacademic hospitals were included.

Fig. 1. PRISMA flowchart.

Search Strategy

PubMed, Google Scholar, SciELO, and Scopus searches were conducted by 2 independent authors (C.M.C.-E. and M.M.-Y.) to identify articles published between 1958 and March 2020 that fit the inclusion criteria. Studies were screened for the availability of full texts. The following keywords were used: (["oral cavity cancer" OR "oral cavity squamous cell carcinoma" OR "machine learning" OR "oral cancer" AND "prognosis"]). Wherever applicable, a manual review of relevant articles referenced was carried out to identify studies missed using the search strategy (Fig. 1). Finally, a critical analysis of the selected studies was performed (Table 1). Ethics Committee approval was not required for this review.

Assessment of Evidence

Data extraction was done in duplicate to avoid errors. Evidence level of studies included was appraised using the Oxford Center for Evidence-Based Medicine (OCEBM) Levels of Evidence [32]. The risk of bias in individual cohort studies was assessed according to the risk of bias in nonrandomized studies of interventions tool (ROBIN-I) [33]. Missing data among studies included were revised and summarized.

Results

A total of 81 manuscripts were reviewed, and 46 studies met the inclusion criteria. Of those, 38 were excluded for the following reasons: use of a classical statistical method ($N = 16$), nonspecific for OCSCC ($N = 15$), and not being related to OCSCC survival ($N = 7$). In total, 8 studies were included in the final analysis. All eight were studied and analyzed; the information is summarized in Table 1 [34–41].

The demographic data available for the included studies are summarized in Table 1. There was a high variation in the number of patients included in each study, ranging between 31 and 33,065 patients [34–41]. In all the studies, the strategy followed corresponded to a supervised ML algorithm approach, and the type of algorithms used were dynamic Bayesian network, artificial neural networks (ANN), support vector machine (SVM), logistic regres-

Table 1. Data from studies included

References	Country	Design	Patients characteristics	Algorithm/ method	Study objective	Goal	Variables/predictors included in training	Performance	Level of evidence
Exarchos et al. [34]	Greece/ Italy/ Spain	Decision support system Supervised ML algorithm	86 patients (characteristics were not described)	DBN, NB, ANN, SVM, DT, RF	Multilevel and multiparametric approach in order to model the growth and progression of OCSCC after remission	Evaluate progression	Baseline data analysis where clinical, imaging, tissue genomic, and blood genomic data were employed in order to predict a potential disease relapse Disease evolution monitoring aiming to capture the progression of the disease based on gene expression data acquired from circulating blood cells and subsequently exploit this information in order to predict if and roughly when a relapse is more likely to appear	DBN accuracy of 86%	3
Exarchos et al. [35]	Greece	Supervised ML algorithm Retrospective cohort	41 patients (characteristics were not described)	BN, ANN, SVM, DT, RF	Propose an orchestrated approach in order to systematically study and analyze the multifactorial basis of oral cancer evolvement	Recurrence prognosis	Clinical, radiological, and genomic data	The proposed approach resulted in perfect discrimination (accuracy: 100%) between patients with and without disease recurrence	3
Chang et al. [36]	Malaysia	Supervised ML algorithm Retrospective cohort	31 oral cancer patients	ANN, SVM – supervised versus LR	Hybrid feature selection and machine learning methods in oral cancer prognosis based on the parameters of the correlation of clinicopathologic and genomic markers	3-years prognosis	Social demographic data (risk factors, ethnicity, age, occupation, marital status, and others) Clinical data (type of lesion, size of lesion, primary site, clinical neck node, etc.) Pathological data (pathological TNM, neck node metastasis, bone invasion, tumor thickness) Genomic variables = p53 & p63	The hybrid model resulted in the best accuracy (accuracy = 93.81%, AUC = 0.90) with the selected features of drink, invasion, and p63	3

Table 1 (continued)

References	Country	Design	Patients characteristics	Algorithm/method	Study objective	Goal	Variables/predictors included in training	Performance	Level of evidence
Alabi et al. [37]	Finland/Brazil	Supervised ML algorithm Retrospective cohort	Early oral tongue cancer 124 = T1 187 = T2 311 = N0 & M0 89 recurrence	ANN – supervised versus LR	Estimate the risk of locoregional recurrence in early-stage OTSCC through a web-based machine learning application	Recurrence prognosis	Age, gender, T stage (AJCC 7th), and WHO histological: Grade, tumor budding, DOI, worst pattern of invasion. Lymphocytic host response and perineural invasion. Follow-up, disease-free time	ANN overall accuracy = 92.7% LR overall accuracy: 86.5%	3
Alabi et al. [38]	Finland/Brazil	Supervised ML algorithm Retrospective cohort	Early oral tongue cancer 124 = T1 187 = T2 311 = N0 & M0 89 recurrence	SVM, NB, BDT, DF, PFI	Compared the performance of 4 machine learning algorithms for predicting the risk of locoregional recurrences in patients with OTSCC	Recurrence prognosis	Age, gender, stage, grade, tumor budding, DOI, worst pattern of invasion, lymphocytic host response, perineural invasion and treatment given	The SVM showed an accuracy of 68% and F1 score of 0.63, NB an accuracy of 70% and F1 score of 0.64, BDT an accuracy of 81% and F1 score of 0.78, and DF an accuracy of 78% and F1 score of 0.70	3
Bur et al. [39]	USA	Supervised ML algorithm Retrospective cohort	1,961 cases from NCDB and 71 cases from author institution T1-2N0M0	LR, DF, SVM, and GBMLA	To develop and validate an algorithm to predict OLNM in clinically node-negative OCSCC patients using machine learning. To compare algorithm performance to a model based on tumor DOI	OLNM prediction	Age, sex, race, ethnicity, primary tumor site, histology, grade, and DOI	The best classification performance was achieved with a DF algorithm (AUC = 0.840)	3
Mermod et al. [40]	Switzerland/Australia	Supervised ML algorithm Retrospective cohort	56 patients in the training cohort (T1–T4) and 112 in the external validation cohort (T–T2)	RF, SVM, LR	Develop, compare, and validate several machines learning models to predict OLNM in clinically N0 (cN0) OCSCC combining markers of lymphangiogenesis and angiogenesis with clinicopathological features	OLNM prediction	Mean intratumoral and peritumoral CD31 MVD, intratumoral and peritumoral PROX1 ND, sex, age, T classification, tumor thickness, lymphovascular space invasion, and perineural spread. The TNM classification was done according to the seventh edition of the AJCC staging system for oral cancer	According to the validation cohort, authors estimated that their model would have reduced the risk of overtreatment of the neck from 82% to only 9%. However, 4% of patients would still present with overlooked OLNM	3

Table 1 (continued)

References	Country	Design	Patients characteristics	Algorithm/method	Study objective	Goal	Variables/predictors included in training	Performance	Level of evidence
Karadaghy et al. [41]	USA	Supervised ML algorithm Retrospective cohort	33,065 patients from the NCDB	DF, ANN, LR, DJ	To develop a prediction model using machine learning for 5-year overall survival among patients with oral squamous cell carcinoma and compare this model with a prediction Model created from the TNM (clinical and pathologic stage	Overall survival prediction	Patient characteristics include age, sex, race/ethnicity, and comorbid disease Tumor characteristics collected include the T, N, and M score and stage group determined both clinically and pathologically. Additional analyzed tumor variables include tumor grade, extracapsular spread, and perineural invasion	The calculated AUC for this machine learning model was 0.80 (95% CI, 0.79–0.81), accuracy was 71%, precision was 71%, and recall was 68% In comparison, the calculated AUC of the TNM staging system was 0.68 (95% CI, 0.67–0.70), accuracy was 65%, precision was 69%, and recall was 52%	3

OTSCC, oral tongue squamous cell carcinoma; DF, decision forest; MVD, microvascular densities; DJ, decision jungle; PFI, permutation feature importance; BDT, boosted decision tree; DT, decision tree; BN, Bayesian network; GBMLA, gradient boosting machine learning architecture; NB, naive Bayes; ND, nuclear densities; RF, random forest; DBN, dynamic Bayesian network; NCDB, National Cancer Database.

sion (LR), decision forest, random forest, gradient boosting machine learning architecture, decision jungle, permutation feature importance, boosted decision tree, naive Bayes, decision tree, and Bayesian network [34–41]. Except in one case [37], all the studies corresponded to a comparison among different ML algorithms.

The main targets described in the included studies corresponded to the evaluation of recurrence prognosis in 3 studies [35, 37, 38]; OLNM prediction in 2 studies [39, 40]; risk of progression, 3-year prognosis, and overall survival prediction in one study, respectively [34, 36, 41]. The most commonly described metric was accuracy, which ranged between 68% and 100% [34–41]. Other metrics described were the AUC, F1 score, precision, and recall. Metrics from studies included related to the AUC, sensitivity (SE), specificity (SP), positive predictive value (PPV), and negative predictive value (NPV) were described in online supplementary Table 1 (see www.karger.com/doi/10.1159/000520672 for all online suppl. material).

According to the Oxford Center for Evidence-Based Medicine grading system, all studies received a grading of 3. The risk of bias according to ROBIN-I is shown in Table 2. Missing data are summarized in online supplementary Table 2.

Discussion

OCSCC represents a significant health problem globally [42, 43]. Despite the wide range of prognostic indicators described, the absolute risk estimated for individual patients remains understudied and is not commonly applied when counseling patients [44, 45]. Clinicians generally rely on the tumor, node, metastasis (TNM) classification to convey the severity of a cancer diagnosis and its prognosis [46]. The increasing use of computer support in hospitals around the world and the increasing availability of large national databases create an outstanding opportunity to explore possibilities of prediction techniques beyond traditional statistics [47].

Evidence about the ML Algorithm as Computer Assistant Tool to Evaluate the OCSCC Prognosis

ML algorithms offer different results that can be categorized as variables; usually these variables can be the *AUC* that measures the ability of a binary classifier to distinguish between classes and is used as a summary of the ROC curve. *SE*, defined as a measure of the proportion of current positive cases that got predicted as positive (true

Table 2. Risk of bias in individual cohort studies (ROBINS-I)

Author	Bias due to confounding	Bias in selection of participants into the study	Bias in classification of interventions	Bias due to deviations from intended intervention	Bias due to missing data	Bias in measurement of outcomes	Bias in selection of the reported results
Exarchos et al. [34]	Low risk	Moderate risk	Moderate risk	Low risk	Moderate risk	Low risk	Low risk
Exarchos et al. [35]	Low risk	Moderate risk	Low risk	Low risk	Moderate risk	Low risk	Low risk
Chang et al. [36]	Low risk	Moderate risk	Moderate risk	Low risk	Moderate risk	Low risk	Low risk
Alabi et al. [37]	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Alabi et al. [38]	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Bur et al. [39]	Low risk	Low risk	Moderate risk	Low risk	Moderate risk	Low risk	Low risk
Mermod et al. [40]	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Karadaghy et al. [41]	Low risk	Low risk	Moderate risk	Low risk	Moderate risk	Low risk	Low risk

positive). *SP*, defined as the proportion of current negative cases, which got predicted as negative (true negative). *PPV*, which can be defined as the proportion of predicted positives cases which are current positives and reflects the probability of a predicted positive to be a true positive. *NPV*, which can be defined as the proportion of predicted negative cases which are real negatives and reflects the probability that a predicted negative case can be a true negative. However, despite these concepts are well known for almost every doctor, it is sometimes difficult to translate these concepts to the clinical scenario and obtain valid conclusions for clinical purposes.

We identified the absence of values like PPV or NPV in some papers, something related directly with the scientific nature of these manuscripts because metrics like sensibility and SP are the most used in the clinical scenario, despite metrics like PPV or NPV besides being similar, are more useful to explain prognosis to our patients. Also, it is important to know that in ML metrics, precision can correspond to PPV, and recall can correspond to SE (online suppl. Table 1).

The first approach to ML in OCSCC was published by Exarchos et al. [34], who compared different algorithms to evaluate the risk of progression among 86 patients affected by OCSCC based on baseline clinical data and disease evolution monitoring data. They reported highest accuracy of 86% using the dynamic Bayesian network [34]. In another paper published by Exarchos et al. [34] to evaluate the risk of recurrence in 41 patients affected by OCSCC, clinical, radiological, and genomic data were used to validate and train the models (Bayesian network, ANN, SVM, decision tree, and random forest). The proposed approach resulted in perfect discrimination (accuracy: 100%) between patients with and without disease

recurrence [35]. A similar study was published by Chang et al. [36] evaluating 3-year survival prognosis. The authors developed a hybrid feature selection model and compared it with ML methods, both based on the correlation of clinicopathologic and genomic markers. They reported that the hybrid model resulted in the best accuracy (accuracy = 93.81%, AUC = 0.90) with the selected features of tissue invasion and p63 being the most relevant prognostic markers [36].

In 2019, at least 5 papers related to the use of ML in OCSCC prognosis were published. Alabi et al. [37, 38] published 2 studies in which they summarized data from Finland and Brazil to estimate the risk of locoregional recurrence in early-stage SCC of the tongue. Both papers are in agreement with the initial large sample size studies using the same dataset [37, 38]. In the first paper, the authors compared the use of an ANN versus LR (overall accuracy was 92.7% vs. 86.5%) using a web-based application [37]. In the second paper, the authors compared different types of algorithms with the SVM and naive Bayes, which reported the highest accuracy of 68% and 70%, respectively [38]. Bur et al. [39] published a study including 1,961 patients from the National Cancer Database in the USA that included clinical and histopathological variables like age, sex, race, ethnicity, primary tumor site, histology, grade, and DOI in an attempt to develop and validate an algorithm to predict OLNM in clinically node-negative OCSCC. They reported that the decision forest was the best classifier, achieving an AUC of 0.840. The most relevant finding of this study was the fact that for the single-institution data, the predictive performance of ML exceeded that of the DOI model (AUC = 0.657, $p = 0.007$), and that when compared to the DOI model, ML reduced the number of recommended neck dissections,

while simultaneously improving SE and SP [39]. Mermod et al. [40] also tried to develop, compare, and validate a different ML algorithm to estimate the risk of OLNM, by combining markers of lymphangiogenesis and angiogenesis with clinicopathological features. According to the validation cohort, the authors estimated that their model would have reduced the risk of overtreating the neck from 82% to only 9%. However, the authors do highlight that in 4% of patients, OLNM would still have been overlooked [40].

In the largest published study including 33,065 patients from the National Cancer Database, Karadaghy et al. [41] intended to develop a model to predict the 5-year overall survival in patients with OCSCC. Variables included patient characteristics such as age, sex, race/ethnicity, and comorbid disease; tumor characteristics included T, N, and M scores and staging as determined both clinically and pathologically as well as some additional tumor variables such as tumor grade, extracapsular spread, and perineural invasion. After training the algorithms, the authors compared the accuracy of the model against the TNM staging. The reported AUC for this ML model was 0.80 (95% CI: 0.79–0.81), with an accuracy of 71%, precision of 71%, and recall of 68%. In comparison, the AUC of the TNM staging system was 0.68 (95% CI: 0.67–0.70), with an accuracy of 65%, a precision of 69%, and a recall of 52% [41].

Implications for Clinical Practice

Currently, the American Joint Commission of Cancer (AJCC) 8th staging system is the most widely used system for assessment of prognosis in OCSCC patients. Therefore, some authors propose the use of nomograms enable to visualize the prognostic strength of various relevant factors in a single model, with a higher accuracy regarding survival prediction than conventional TNM staging system or an individual molecular biomarker [48–51].

However, previous authors considered that traditional statistical methods such as Cox proportional hazards regression, LR, and Kaplan-Meier may slow the progress of prediction models [41]. This assumption is illustrated by the inability of the aforementioned methods to handle medical data with high variability, nonlinear interactions, heterogeneous distributions or learn new patterns from the data generated prospectively to modify and improve prognosis accuracy [52, 53]. By contrast, ML may be a better instrument to handle large datasets with complex, nonlinear, and heterogeneous distributions [52–54]. This is a situation commonly found when we try to analyze large cancer databases, including different types of vari-

ables [55]. Another common difference between traditional models studies and ML studies, correspond to the assessment of the data including only sample analyzed without the inclusion of new data within the algorithm.

At this early point of ML research, the real impact of these techniques on OCSCC prognostication is difficult to determine. Also, validating the role of ML algorithms in OCSCC is difficult at this point as there are unanswered questions relating to, e.g., whether one can trust ML algorithms to predict prognosis, whether data collection was good, which one is the best modeling algorithm, and whether it is possible to agree on the use of this technology in a multidisciplinary cancer committee setting. One would also need to convince doctors involved in the treatment of OCSCC about the benefits of this technique. All these questions are relevant to assess the real impact of ML in a real-life clinical setting. However, except for some web-based applications used in a research environment, there is currently no widely available implementation in a real clinical setting.

Despite this, the predictive ability of ML and its ability to improve recently developed methods such as DOI to predict the risk of OLNM or even to improve prognostication of the current TNM system, by combining these data with demographic, clinical, and histopathological data, and other variables can be accepted from the reported literature [39, 41]. Use of ML-based computer prognosis support may be translated into a more precise prognosis of survival and reduce surgical or treatment morbidity, thereby improving patients' quality of life and permitting more accurate counseling. In addition, the development of such algorithms is affordable, especially with the support of a great open-source community providing a host of solutions to develop the software. ML may also permit the real possibility to fine-tune and adapt prognostic tools to suit demographic, clinical, and histopathological data of individual countries or regions, and even take into consideration resource constraints.

Limitations and Future Challenges Related to the Application of ML in OCSCC

As we have noted, almost all the studies till date have been designed to be a proof of concept about the feasibility of building ML-based cancer prognostic tool. There remain significant challenges in applying ML techniques in the prediction of the prognosis of OCSCC.

Some of these challenges include the relatively small number of patients and data available, and the retrospective nature of almost all the studies, increasing the risk of suboptimal performance due to overfitting problems. It

is necessary to apply regularization methods such as ridge and lasso or L1 and L2, among others, to combat this. Imbalanced patient cohorts can be another limitation, e.g., in some high-mortality cancers like OCSCC, it is common to find fewer survivors in the study group. This might affect the performance of the algorithm because of the imbalanced data used to train. The high risk of missing data from highly completed datasets is also a challenge. To overcome this problem, a data imputation method based on known data can be used, as proposed by Rendleman et al. [55] who described a multivariate imputation by the chained equations (MICE) model that works with the assumption that the missing data are missing in a random fashion.

Another challenge is the need for a more generic consensus in the best ML training algorithm because of the difficulty in evaluating model accuracy among different studies. One solution is to start building cancer patient databases for prognosis analysis [56–60]. However, specific infrastructure is needed for data storage to build, develop, and train ML models. Such endeavors also need to consider privacy requirements of healthcare data, and the buy-in of political administrations, the research community, and personal awareness [61]. Finally, there is a global need for computational researchers who have expertise in biomedical research and ML to improve research in this area. Finally, we need to highlight the little available evidence about ML as a computer assistant approach across the literature as a significant limitation of our study.

Conclusion

ML has the potential to significantly advance research in the field of OCSCC. Advantages are related to the use and training of ML models because of their capability to continue training continuously when more data become available. Future ML research will allow us to improve and democratize the application of algorithms to improve the prediction of cancer prognosis and its management worldwide.

Statement of Ethics

Ethics Committee approval for systematic reviews was not required in any of the participant institutions.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Funding Sources

No funding was received for this study.

Author Contributions

1. Carlos M. Chiesa-Estomba: A, B, C, D.
2. Manuel Graña: A, C.
3. Alfonso Medela: A.
4. Jon A. Sistiaga-Suarez: A.
5. Jerome R. Lechien: A.
6. Christian Calvo-Heriquez: A.
7. Miguel Mayo-Yanez: B.
8. Luigi Angelo Vaira: B.
9. Alberto Grammatica: B.
10. Giovanni Cammaroto: D.
11. Tareck Ayad: D.
12. Johannes J. Fagan: D.
    A Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND
    B Drafting the work or revising it critically for important intellectual content; AND
    C Final approval of the version to be published; AND
    D Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Data Availability Statement

All data generated or analyzed during this study are included in this article (and/or) its online supplementary material. Further inquiries can be directed to the corresponding author.

References

1 World Health Organization. Oral health. 2018. Available from: https://gco.iarc.fr/today/data/factsheets/cancers/1-Lip-oral-cavity-fact-sheet.pdf Accessed 2020 May 15.
2 Ferlay JEM, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. Int J Cancer. 2019;144:1941–53.
3 NCCN. Head and neck cancers: NCCN clinical practice guidelines in oncology. [updated 2018 Feb 15]. Available from: https://oncolife.com.ua/doc/nccn/Head_and_Neck_Cancers.pdf Accessed 2020 Mar 2.
4 Huang TH, Li KY, Choi WS. Lymph node ratio as prognostic variable in oral squamous cell carcinomas: systematic review and meta-analysis. Oral Oncol. 2019;89:133–43.

5 Gharat SA, Momin M, Bhavsar C. Oral squamous cell carcinoma: current treatment strategies and nanotechnology-based approaches for prevention and therapy. Crit Rev Ther Drug Carrier Syst. 2016;33:363–400.

6 Torrecillas V, Shepherd HM, Francis S, Buchmann LO, Monroe MM, Lloyd S, et al. Adjuvant radiation for T1-2N1 oral cavity cancer survival outcomes and utilization treatment trends: analysis of the SEER database. Oral Oncol. 2018;85:1–7.

7 Camisasca DR, Honorato J, Bernardo V, da Silva LE, da Fonseca EC, de Faria PA, et al. Expression of Bcl-2 family proteins and associated clinicopathologic factors predict survival outcome in patients with oral squamous cell carcinoma. Oral Oncol. 2009;45:225–33.

8 Mermod M, Bongiovanni M, Petrova TV, Dubikovskaya EA, Simon C, Tolstonog G, et al. Prediction of occult lymph node metastasis in squamous cell carcinoma of the oral cavity and the oropharynx using peritumoral Prospero homeobox protein 1 lymphatic nuclear quantification. Head Neck. 2016;38:1407–15.

9 Ferlito A, Rinaldo A, Devaney KO, MacLennan K, Myers JN, Petruzzelli GJ, et al. Prognostic significance of microscopic and macroscopic extracapsular spread from metastatic tumor in the cervical lymph nodes. Oral Oncol. 2002;38:747–51.

10 Psychogios G, Mantsopoulos K, Bohr C, Koch M, Zenk J, Iro H. Incidence of occult cervical metastasis in head and neck carcinomas: development over time. J Surg Oncol. 2013;107:384–7.

11 Goldstein DP, Ringash J, Bissada E, Jaquet Y, Irish J, Chepeha D, et al. Scoping review of the literature on shoulder impairments and disability after neck dissection. Head Neck. 2014;36:299–308.

12 Fukano H, Matsuura H, Hasegawa Y, Nakamura S. Depth of invasion as a predictive factor for cervical lymph node metastasis in tongue carcinoma. Head Neck. 1997;19:205–10.

13 Girod A, Mosseri V, Jouffroy T, Point D, Rodriguez J. Women and squamous cell carcinomas of the oral cavity and oropharynx: is there something new? J Oral Maxillofac Surg. 2009;67:1914–20.

14 Thiagarajan S, Nair S, Nair D, Chaturvedi P, Kane SV, Agarwal JP, et al. Predictors of prognosis for squamous cell carcinoma of oral tongue. J Surg Oncol. 2014;109:639–44.

15 Liu F, Chen F, Huang J, Yan L, Liu F, Wu J, et al. Prospective study on factors affecting the prognosis of oral cancer in a Chinese population. Oncotarget. 2017;8:4352–9.

16 Jadhav KB, Gupta N. Clinicopathological prognostic implicators of oral squamous cell carcinoma: need to understand and revise. N Am J Med Sci. 2013;5:671–9.

17 Bur AM, Lin A, Weinstein GS. Adjuvant radiotherapy for early head and neck squamous cell carcinoma with perineural invasion: a systematic review. Head Neck. 2016;38 Suppl 1:E2350–7.

18 Sparano A, Weinstein G, Chalian A, Yodul M, Weber R. Multivariate predictors of occult neck metastasis in early oral tongue cancer. Otolaryngol Head Neck Surg. 2004;131:472–6.

19 Arora A, Husain N, Bansal A, Neyaz A, Jaiswal R, Jain K, et al. Development of a new outcome prediction model in early-stage squamous cell carcinoma of the oral cavity based on histopathologic parameters with multivariate analysis: the Aditi-Nuzhat Lymph-node Prediction Score (ANLPS) system. Am J Surg Pathol. 2017;41:950–60.

20 Sawant SS, Dongre H, Ahire C, Sharma S, Kannan S, Mahadik S, et al. A nomogram for predicting the risk of neck node metastasis in pathologically node-negative oral cavity carcinoma. Oral Dis. 2017;23:1087–98.

21 De Silva RK, Siriwardena BSMS, Samaranayaka A, Abeyasinghe WAMUL, Tilakaratne WM. A model to predict nodal metastasis in patients with oral squamous cell carcinoma. PLoS One. 2018;13:e0201755.

22 Bur AM, Shew M, New J. Artificial intelligence for the otolaryngologist: a state of the art review. Otolaryngol Head Neck Surg. 2019;160:603.

23 Obermeyer Z, Emanuel EJ. Predicting the future: big data, machine learning, and clinical medicine. N Engl J Med. 2016;375:1216–9.

24 Michie D, Spiegelhalter DJ, Taylor CC. Machine learning, neural and statistical classification. New York: Ellis Horwood; 1994.

25 van Hooff SR, Leusink FK, Roepman P, Baatenburg de Jong RJ, Speel EJ, van den Brekel MW, et al. Validation of a gene expression signature for assessment of lymph node metastasis in oral squamous cell carcinoma. J Clin Oncol. 2012;30:4104–10.

26 de Oliveira MV, Pereira Gomes EP, Pereira CS, de Souza LR, Barros LO, Mendes DC, et al. Prognostic value of microvessel density and p53 expression on the locoregional metastasis and survival of the patients with head and neck squamous cell carcinoma. Appl Immunohistochem Mol Morphol. 2013;21:444–51.

27 Mermod M, Bongiovanni M, Petrova T, Goun E, Simon C, Tolstonog G, et al. Prediction of occult lymph node metastasis in head and neck cancer with CD31 vessel quantification. Otolaryngol Head Neck Surg. 2019;160:277–83.

28 Beasley NJ, Prevo R, Banerji S, Leek RD, Moore J, van Trappen P, et al. Intratumoral lymphangiogenesis and lymph node metastasis in head and neck cancer. Cancer Res. 2002;62:1315–20.

29 Franchi A, Gallo O, Massi D, Baroni G, Santucci M. Tumor lymphangiogenesis in head and neck squamous cell carcinoma: a morphometric study with clinical correlations. Cancer. 2004;101:973–8.

30 Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009;339(6):b2535.

31 Thompson M, Tiwari A, Fu R, Moe E, Buckley DI. A framework to facilitate the use of systematic reviews and meta-analyses in the design of primary research studies [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2012.

32 Howick J, Chalmers I, Glasziou P, Greenhalgh T, Heneghan C, Howick J. The 2011 Oxford CEBM levels of evidence (introductory document). 2011. Available from: https://www.cebm.net/index.aspx?o=5653 Accessed 2020 May 12.

33 Sterne JA, Hernán MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in nonrandomised studies of interventions. BMJ. 2016;355:i4919.

34 Exarchos KP, Goletsis Y, Fotiadis DI. A multiscale and multiparametric approach for modeling the progression of oral cancer. BMC Med Inform Decis Mak. 2012;12:136.

35 Exarchos KP, Goletsis Y, Fotiadis DI. Multiparametric decision support system for the prediction of oral cancer reoccurrence. IEEE Trans Inf Technol Biomed. 2012;16:1127–34.

36 Chang SW, Abdul-Kareem S, Merican AF, Zain RB. Oral cancer prognosis based on clinicopathologic and genomic markers using a hybrid of feature selection and machine learning methods. BMC Bioinformatics. 2013;14:170.

37 Alabi RO, Elmusrati M, Sawazaki-Calone I, Kowalski LP, Haglund C, Coletta RD, et al. Machine learning application for prediction of locoregional recurrences in early oral tongue cancer: a web-based prognostic tool. Virchows Arch. 2019;475:489–97.

38 Alabi RO, Elmusrati M, Sawazaki-Calone I, Kowalski LP, Haglund C, Coletta RD, et al. Comparison of supervised machine learning classification techniques in prediction of locoregional recurrences in early oral tongue cancer. Int J Med Inform. 2020;136:104068.

39 Bur AM, Holcomb A, Goodwin S, Woodroof J, Karadaghy O, Shnayder Y, et al. Machine learning to predict occult nodal metastasis in early oral squamous cell carcinoma. Oral Oncol. 2019 May;92:20–5.

40 Mermod M, Jourdan EF, Gupta R, Bongiovanni M, Tolstonog G, Simon C, et al. Development and validation of a multivariable prediction model for the identification of occult lymph node metastasis in oral squamous cell carcinoma. Head Neck. 2020;42:1811–20.

41 Karadaghy OA, Shew M, New J, Bur AM. 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 May 2;145:1115–20.

42 Funk GF, Karnell LH, Robinson RA, Zhen WK, Trask DK, Hoffman HT. Presentation, treatment, and outcome of oral cavity cancer: a national cancer data base report. Head Neck. 2002;24:165–80.

43 Rivera C. Essentials of oral cancer. Int J Clin Exp Pathol. 2015;8:11884–94.

44 Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J. 2014;35:1925–31.

45 Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ. 2009;338: b606.

46 Piccirillo JF. Purposes, problems, and proposals for progress in cancer staging. Arch Otolaryngol Head Neck Surg. 1995;121:145–9.

47 Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med. 2019; 25:24–9.

48 Mattavelli D, Lombardi D, Missale F, Calza S, Battocchio S, Paderno A, et al. Prognostic nomograms in oral squamous cell carcinoma: the negative impact of low neutrophil to lymphocyte ratio. Front Oncol. 2019;9:339.

49 Wang F, Zhang H, Wen J, Zhou J, Liu Y, Cheng B, et al. Nomograms forecasting long-term overall and cancer-specific survival of patients with oral squamous cell carcinoma. Cancer Med. 2018;7(4):943–52.

50 Kao HK, Lofstrand J, Loh CY, Lao WWK, Yi JS, Chang YL, et al. Nomogram based on albumin and neutrophil-to-lymphocyte ratio for predicting the prognosis of patients with oral cavity squamous cell carcinoma. Sci Rep. 2018;8(1):13081.

51 Montero PH, Yu C, Palmer FL, Patel PD, Ganly I, Shah JP, et al. Nomograms for preoperative prediction of prognosis in patients with oral cavity squamous cell carcinoma. Cancer. 2014;120(2):214–21.

52 Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. Cancer Inform. 2007;2:59–77.

53 Wang G, Lam KM, Deng Z, Choi KS. Prediction of mortality after radical cystectomy for bladder cancer by machine learning techniques. Comput Biol Med. 2015;63:124–32.

54 McCarthy JF, Marx KA, Hoffman PE, Gee AG, O'Neil P, Ujwal ML, et al. Applications of machine learning and high-dimensional visualization in cancer detection, diagnosis, and management. Ann N Y Acad Sci. 2004;1020: 239–62.

55 Rendleman MC, Buatti JM, Braun TA, Smith BJ, Nwakama C, Beichel RR, et al. Machine learning with the TCGA-HNSC dataset: Improving usability by addressing inconsistency, sparsity, and high-dimensionality. BMC Bioinformatics. 2019;20:339.

56 Goswami CP, Nakshatri H. PROGgene: gene expression based survival analysis web application for multiple cancers. J Clin Bioinforma. 2013;3:22.

57 Anaya J. OncoLnc: linking TCGA survival data to mRNAs, miRNAs, and lncRNAs. Peer J Comput Sci. 2016;2.

58 Elfilali A, Lair S, Verbeke C, La RP, Radvanyi F, Barillot E. ITTACA: a new data base for integrated tumor transcriptome array and clinical data analysis. Nucleic Acids Res. 2006; 34:D613–6.

59 Wang Q, Xie L, Dang Y, Sun X, Xie T, Guo J, et al. A web server to evaluate the prognostic value of genes in leiomyosarcoma. Front Oncol. 2019;9:190.

60 Wang Q, Wang F, Lv J, Xin J, Xie L, Zhu W, et al. Interactive online consensus survival tool for esophageal squamous cell carcinoma prognosis analysis. Oncol Lett. 2019;18:1199–206.

61 Wang F, Wang Q, Li N, Ge L, Yang M, An Y, et al. An interactive online consensus survival tool for uveal melanoma prognosis analysis. Mol Carcinog. 2020;59:56.