Machine learning in oral squamous cell carcinoma: current status, clinical concerns and prospects for future - A systematic review

Author(s): Alabi, Rasheed Omobolaji; Youssef, Omar; Pirinen, Matti; Elmusrati, Mohammed; Mäkitie, Antti A.; Leivo, Ilmo; Almangush, Alhadi

Title: Machine learning in oral squamous cell carcinoma: current status, clinical concerns and prospects for future - A systematic review

Year: 2021

Version: Author accepted manuscript

Please cite the original version:

Alabi, R. O., Youssef, O., Pirinen, M., Elmusrati, M., Mäkitie, A. A., Leivo, I. & Almangush, A. (2021). Machine learning in oral squamous cell carcinoma: current status, clinical concerns and prospects for future - A systematic review. *Artificial intelligence in medicine* 115. https://doi.org/10.1016/j.artmed.2021.102060

Machine learning in oral squamous cell carcinoma: current status, clinical concerns and prospects for future

Rasheed Omobolaji Alabi M.Sc [a], Matti Pirinen [d], Mohammed Elmusrati D.Sc [a], Antti A. Mäkitie MD, PhD [c,e], Ilmo Leivo MD, PhD [f*], Alhadi Almangush DDS, PhD [b,c,f,g*]

[a] Department of Industrial Digitalization, School of Technology and Innovations, University of Vaasa, Vaasa, Finland.

[b] Department of Pathology, University of Helsinki, Helsinki, Finland.

[c] Research Program in Systems Oncology, Faculty of Medicine, University of Helsinki, Helsinki, Finland.

[d] Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland.
Department of Public Health, University of Helsinki, Helsinki, Finland.
Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland.

[e] Department of Otorhinolaryngology – Head and Neck Surgery, University of Helsinki and Helsinki University Hospital, Helsinki, Finland.
Division of Ear, Nose and Throat Diseases, Department of Clinical Sciences, Intervention and Technology, Karolinska Institutet and Karolinska University Hospital, Stockholm, Sweden.

[f] University of Turku, Institute of Biomedicine, Pathology, Turku, Finland.

[g] Faculty of Dentistry, University of Misurata, Misurata, Libya.

***The last two authors have equal contributions.**

Manuscript text count: 2,959

Corresponding Author: Rasheed Omobolaji Alabi
Department of Industrial Digitalization, School of Technology and Innovations, University of Vaasa, Vaasa, Finland. **E-mail address:** rasheed.alabi@student.uwasa.fi

Disclosure: The authors declare no conflicts of interest.

Abstract

Importance: Oral cancer can show heterogenous patterns of behavior. For proper and effective management of oral cancer, early diagnosis and prognosis are important. To achieve this, artificial intelligence (AI) or its subfield, machine learning, has been touted for its potential to revolutionize cancer management through improved diagnostic precision and prediction of outcomes. Yet, to date, it has made only few contributions to actual medical practice or patient care. **Objectives:** This study provides a state of art the review of diagnostic and prognostic roles of machine learning in oral squamous cell carcinoma (OSCC) and also highlights some of the limitations and concerns of clinicians towards the implementation of these models into daily clinical practice. **Design:** We searched OvidMedline, PubMed, Scopus, Web of Science, and Institute of Electrical and Electronics Engineers (IEEE) databases for articles that used machine learning for diagnostic or prognostic purposes of OSCC. We used the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) in the searching and screening processes. **Main outcomes and measures:** The clinical concerns for the integration of machine learning models for actual daily practice in oral tongue cancer were identified. **Results:** A total of 41 studies were reported to have used machine learning to analyse of OSCC. The majority of these studies used support vector machine (SVM) and artificial neural network (ANN) algorithms as machine learning techniques. Their specificity ranged from 0.57 to 1.00, sensitivity from 0.70 to 1.00, and accuracy from 63.4% to 100.0% in these studies. The main limitations and concerns were a lack of proper understanding of the used machine learning models, inability to interpret which aspect of the data contributes to the result, concern about models possibly rendering the clinicians less important in patient management decisions, and privacy violation. **Conclusion:** The accumulated evidence indicates that machine learning models have a great potential in improving survival of OSCC patients. Therefore, it is important that the concerns of the clinicians are taken into consideration in the development of machine learning models. This would allow for a seamless integration of these models into the daily clinical practice.

KEYWORDS: Machine learning; Oral squamous cell carcinoma; Systematic review; explainable AI

1. Introduction

Oral cancer is an aggressive disease characterized by a low average survival rate [1]. Developments in treatment modalities in the domains of both oncology and surgery have only contributed to a rather limited improvement in outcome. Therefore, accurate diagnosis and prognosis prediction of cancer, especially at an early stage are important in improving survival rate [2]. The availability of different treatment options for oral cancer requires a proper selection of the treatment on a case-by-case basis.

Despite improved effect of the treatment, individualized patient-specific treatments are mostly lacking. Thus, improvements in diagnostic and prognostic accuracy could significantly assist the clinicians in making informed decisions on treatment. To this end, technical advances in statistics and computer software have led to improved prognostication using multi-factor analysis via conventional logistic and Cox regression models. Similarly, the application of machine learning techniques, a subfield of artificial intelligence (AI), plays a major role in the improved prediction of cancer outcomes. Several studies have reported that machine learning approach is more accurate in prognostication than the traditional statistical analyses [3–7].

Machine learning approach was found to be beneficial in the three aspects that are essential to early diagnosis and prognosis. These are an improved accuracy of cancer susceptibility, recurrence, and survival predictions [2], which improve the survival rates through the effective clinical management of patients [8–14]. Over the coming years, the application of the machine learning approach to clinical research continues to increase due to its feasibility and its many advantages. For instance, our group has used machine learning techniques to predict the locoregional recurrence of oral tongue cancer [15]. Similarly, it has been used to detect oral cancer [16–22], and to predict oral cancer recurrence [23,24], occult node metastasis [25,26], and survival rates of oral cancer [27–30]. Additionally, it has been used for the prognostication of other cancers [31–33] and to predict progression of diseases on the basis of patient records such as from pre-diabetes to type 2 diabetes based on the patients' records [34]. All these applications of machine learning in healthcare are aimed at assisting the doctors in making informed decisions, reducing diagnostics errors, improving and promoting the overall patient health.

Despite numerous studies on the application of machine learning and various intelligent models deployed, the question remains – what are the concerns of clinicians towards the actual implementation of machine learning-based models in clinical settings? These concerns were considered from the limitations, shortcomings, and clinicians' concerns in the published studies regarding the application of machine learning for oral squamous cell carcinoma (OSCC) prognosis. This study, therefore, aims to systematically review the studies on the application of machine learning for diagnosis and prognosis of oral squamous cell carcinoma. OSCC was chosen in this review as it is the most common malignancy of the oral cavity. Also, it constitutes a majority of head and neck squamous cell carcinoma.

2. Methods

2.1. Search protocol. In this study, we systematically retrieved all studies that applied machine learning techniques to oral cancer diagnosis or prognosis. The systematic search included databases of OvidMedline, PubMed, Scopus, Web of Science, and Institute of Electrical and Electronics Engineers (IEEE) from their inception until February 2020. The search approach was developed by combining search keywords: [('oral cancer') AND ('machine learning')]. An additional search was conducted using the search terms: [('oral cancer') AND ('artificial neural network' OR 'ensemble method')]. To minimize the possibility of omission of any study, the reference lists of all the eligible articles were manually searched to ensure that all the relevant studies were duly included. Also, the Preferred

Reporting Items for Systematic Review and Meta-Analysis (PRISMA) was followed in the searching and screening processes (Figure 1) [35].

2.2. Inclusion and exclusion criteria. The eligible studies must have evaluated the diagnostic or prognostic significance of using machine learning algorithms in oral cancer. Invited reviews, review articles, case series, case reports, abstracts, studies on animals, conference papers, editorials, letters to the editors, commentaries, comparative studies, expert views, and general studies on cancer (not specific to oral cancer) were all excluded. Similarly, articles in languages other than English were excluded. Studies that examined machine learning application for normal oral mucosa, oral lesions (without cancer), oral caries, oral mucosa, DNA and RNA microarray genes, proteomics, fluorescence spectroscopy, genetic programming and Fuzzy systems were excluded. The details of the inclusion and exclusion criteria are described in Figure 1.

2.3. Screening. To ensure that all eligible studies were included in this study, a data extraction sheet was used where the studies selected to meet the required criteria for this review. The data extraction process was conducted by two independent reviewers (A.R., & O.Y.). Possible discrepancies were resolved by discussion. A consensus was reached on which studies should be included or excluded after deliberations considering the objectives, and the inclusion and exclusion criteria of the study.

2.3. Data extraction. The extracted parameters from each study included author (s) name, year of publication, country of authors, site of mouth cancer, number of study participants, machine learning algorithms examined in the study, definition of study objective (prognostic or diagnostic), study aim, results, performance metrics (accuracy and/or specificity, or area under receiving operating characteristics (ROC) curve AUC) reported, and conclusion from the study (Table 1). When more than one algorithm was considered in the study, the algorithm with the best performance metrics was extracted, and included in the corresponding column in

Table 1. Similarly, where the results were reported separately for training and validation sets, the reported results for the validation were presented as shown in Table 1. Other important information, such as the limitations of the study and the prognostic significance of the application of the machine learning technique, were noted and summarized in the Discussion section.

2.4. Quality assessment. We used the guidelines for developing and reporting machine learning predictive models to assess the quality of studies that evaluated the application of machine learning in the prognosis of OSCC [36]. We summarized the main guidelines in Table 2. Each point from the guidelines carries a single mark. The threshold was set to be half of the maximum marks. The details of the studies and the final score from these guidelines are given in Table 3.

3. Results

3.1. Results of the database search. The PRISMA flowchart (Figure 1) describes the study selection process. A total of 297 hits were retrieved. After deleting duplicates (N = 150), irrelevant papers (N = 91), and exclusions (N = 15), we found 41 studies eligible to be included in this systematic review as shown in Figure 1 [5, 15–30, 37–60]. The findings of these studies (summarized in Table 1) indicated that the application of machine learning techniques for oral cancer (diagnosis and/or prognosis) could assist the clinicians in making informed decisions regarding diagnostics and prognostic parameters. The results also indicated that these techniques are poised to offer personalized patient care and could improve survival and reduce the death rate associated with oral cancer. In addition, many of these studies mentioned significant limitations for the adoption of such models to actual daily medical practice.

3.2. Characteristics of relevant studies.

All the articles included were published in the English language. Of the 41 included studies, 35 studies considered oral cavity cancer in general [16–30,37,40,41,43,44,46,48,49,52–60], 4 studies focused on oral tongue squamous cell carcinoma [5,15,50,51], while 2 studies considered other sites in addition to oral cavity [38,47]. Furthermore, 19 studies examined the prognostic significance of machine learning applications, 21 studies evaluated the diagnostic significance of machine learning applications, and one study evaluated both (Table 1). Most studies on the application of machine learning techniques in oral cancer were published recently in 2018 and 2019 (N = 24). Over 90% of the data used in the included studies were retrospective in nature. With regards to the origin of relevant articles, 65.8% of the studies were carried out entirely in Asia, 9.6% in Europe, 7.3% in America, and 17.3% of the studies were collaborative efforts from different regions. Furthermore, a total of 4 (9.8%) of the studies used autofluorescence spectral data analysis in addition to the machine learning techniques [38,40,41,52]. Additionally, 18 (43.9%) studies used clinicopathologic or imaging data [5,15,17–21,24,25,27,28,37,45,48,49,57–59]. Also, 2 (4.9%) studies used either clinicopathologic and image [29,56], or clinicopathologic and genomic [43,44], or genomic data only [46,47], or Raman spectral data [50,51]. A single study (2.4%) combined clinical, imaging and genomic data [23]. Similarly, one study (2.4%) used clinical and genomic data [42], while 9 (21.9%) studies used other types of data (combination of risk habits, personal details, and dental attendance, or histopathologic, saliva samples, demographics and histopathologic, pathologic, lesion conditions and histological grade, clinicopathologic and socio-demographic, histologic and brush cytologic parameters, demographics-histopathologic and immunohistochemical).

Most of the included studies considered artificial neural networks (N =12, 29.3%) or support vector machines (N = 14, 34.1%) in their analyses. These two popular algorithms were followed closely by deep convolutional neural networks (N = 11, 26.8%) [17,19,20,46,48,50–52,57–59]. There was also an increase in the application of deep neural network from the year 2017 onwards. In total, 24 (80%) of the studies had the number of cases less than 500. Similarly, most of the cases used for the analysis were extracted from hospital health records (N = 27, 65.8%). Several metrics were reported in these studies to report the performance of these machine learning algorithms. Of the included studies, 13 (31.7%) reported accuracy as their performance metrics [21–23,28,30,37,43,44,48,49,54,59,60]. Also, 13 (31.7%) used sensitivity, specificity and accuracy [5,15,17,18,26,39,42,45,46,50,51,57,58] while 8 (19.5%) studies employed only sensitivity and specificity [16,20,27,38,40,41,52,55] . Four (7.3%) studies reported only specificity and accuracy [24,25,53,56]. A single study (2.4%) considered sensitivity, specificity, accuracy and area under receiving operating characteristic curve (AUC) [19], while 2 (4.9%) studies used only AUC or its mean (MAUC) [29,47].

A total of 30 studies (73.2%) used a shallow machine learning approach while 11(26.8%) employed a deep machine learning approach. Reported specificity in the reported studies ranged from 0.57 to 1.00 [25,27,41] and sensitivity varied between 0.70 and 1 [16, 27]. Similarly, accuracy ranged from 63.4% to 100%. Notably, only 4 (9.8%) of the included studies reported less than 75% performance accuracy of the machine learning model [18,25,30,45]. Some of the concerns were the black-box concern (inability to interpret how the trained machine learning models make the diagnosis or predictions of the patients on a case-by-case basis) [25,61], result and model interpretability (what aspect of the data or the input features led to the prediction) [25,62,63], the amount and quality of the data used in the training [25,30], super-human analogy (assumption that the diagnosis or prognosis from the machine learning algorithm is close to perfect or better than the performance of the clinicians) [62], generalizability of the model (the predictive model can be used outside the data on which it was trained initially) [5,15,25], job-competitor (concerns that the adoption of machine learning model would replace the pathologists), commercial interests (integration of machine learning-based model may actually reduce the revenue of the health systems and consequently of the clinicians) [25], and ethical issues (protecting the privacy of the patients information and defining who will be responsible if the model fails) [25,30].

3.4. Quality assessment of the studies included in the review

The quality of the studies included in this study was scaled from satisfactory to excellent. Most of the studies were generally good (Table 3). Although some of the studies did not properly follow the guidelines provided by Luo et al. (Table 2).

4.0 Discussion

The number of studies that focus on the application of machine learning in oral cancer has increased in recent years. In this systematic review, we examined for the first time the studies published on the application of machine learning in oral cancer management. The evaluated studies considered the use of machine learning to analyze clinicopathologic data, genomic data, combination of clinicopathologic and genomic data, image data, and autofluorescence spectral data. These approaches generated models to assist in clinical decision making [64].

Interestingly, the performance metrics reported in the included studies suggest high performance. Thus, the application of machine learning for oral cancer, as well as in other fields of medicine is not merely science fiction, but is becoming a reality [65]. This finding was corroborated by another study that examined machine learning and its potential applications to genomic studies of the head and neck [66]. Of note, sensitivity, specificity and accuracy have been the widely reported performance metrics. This is because accuracy simply considers correct predictions over all the predictions made by the algorithm. Similarly, specificity measures the proportion of patients that did not have oral cancer and were predicted by the model as non-oral cancer while sensitivity (recall) measures what proportion of patients actually had oral cancer and were identified by the algorithm as having oral cancer.

Using machine learning techniques, a web-based tool has been developed to predict locoregional recurrence [5]. Similarly, machine learning technique was used to automate the diagnosis of oral cancer [49]. Many prognostic factors have been combined together via machine learning techniques for outcome predictions [15,23–30,43,58]. Also, the approach has demonstrated significant accuracy in discriminating between patients with or without oral cancer [16–19,21,22,38,41,47,52,57,59]. In other contexts to enhance effective management of oral cancer, machine learning techniques were used for early-stage detection of precancerous and cancerous lesions [20,40,46,55,60].

Despite the benefits of ensemble machine learning algorithms, support vector machine (SVM) was the most widely used machine learning algorithm for oral cancer diagnosis/prognosis as shown in this systematic review. This was also noted in a study that examined machine learning and its application to genomic data of head and neck cancer [66]. In another study, the support vector machine was concluded to be the most favorable algorithm for predicting survival rate of oral cancer [45]. The support vector machine is frequently used because it is an empirical risk minimizer algorithm. Additionally, it avoids the danger of being trapped in local minima [67]. Thus, it is usually not prone to overfitting, thereby making it capable of producing a good model that can properly capture the complex relationships between the input and output parameters. Of note, the first study that examined the use of artificial intelligence to identify patients at high risks of oral cancer used artificial neural network (ANN) [16]. Consequently, the neural network was also one of the most widely used algorithms. Success recorded from the use of neural network led to its' modification to contain multiple hidden layers. Hence, the name deep neural networks. Deep neural networks are well-positioned to solve most complex problems such as image analysis [68,69]. The application of deep learning technologies to oral cancer diagnosis and prognosis has increased in recent years [19,20,46,48,51,52,57–59].

All the studies included in this systematic review emphasized that machine learning techniques offer an increased precision approach to clinicians by making informed decisions. This further enhances patient-specific treatments and effective management of hospital resources in a timely, efficient and dynamic manner [5,15–17,20,23,25,30,38,70,71]. Despite these potential benefits, the application of machine learning for medical diagnosis and prognosis has made few contributions to actual medical practice or patient care (Figure 2). Several issues are particularly significant from the clinical and ethical viewpoints.

The first and most frequent issue is the black-box concern [25,61,72] (Figure 3). It comes in from two distinct yet interacting perspectives, namely the result and model interpretability concerns [62]. Result interpretability concern entails an inability of the clinicians to explain which aspect of the dataset used in the training led to the predicted result in a particular case. Similarly, model interpretability reflects the clinicians' ability to understand how the algorithm developed the model [25,62]. As the trend in machine learning techniques moves from direct algorithms, such as support vector machine, to ensemble algorithms, and to deep learning, the black-box concern becomes more pronounced. To address this concern, it is pertinent for the machine learning techniques and the corresponding model to be explainable ("explainable model") and transparent [25,30,61,63] (Figure 4). Clinicians should be able to understand, to trust, to explain and to effectively manage the emerging generation of models to be used for clinical decision making. Several terms have been used to describe this concept. These include explainable AI, transparent ML, interpretable ML, and trustworthy AI [73–75].

The second concerns is the misconceptions of the scope of machine learning in medical diagnosis. The notion that machine learning models are super-human or close to perfect is erroneous and misleading. This has led to the fear and predictions that these models in the nearest future could replace the need for professional experience-based consideration in diagnostics and prognostication [76]. The experience of the machine learning experts and the quality of the data used in machine learning analyses play a central role in producing a good model. Therefore, it is necessary that the quality of data used for model training should be the best possible and well-structured to produce a high-quality model [25,30,77].

The third concern relates to the limited amount of data used in the machine learning analyses [5,17,19,23,28,38,43,44,46,55]. Therefore, there is concern for generalizability concern of the developed machine learning model. Performance of the model to be applied for external cases outside the data for which the model was trained, is a subject to be highlighted [5,15,25,29,38]. Thus, for the machine learning model to create sustainable benefits in medical diagnosis, the data infrastructure of healthcare organizations' needs to be improved and the model produced should be externally validated to avoid biases and to enhance generalizability of the model. In the quest to improve the healthcare organizations' data infrastructure, also privacy of patient information and ethical use of the data should also be considered [25,30]. Of note, a generalized model does not mean a super-human model [62], which is a concern amongst certain clinicians. Rather, it means that the inherent bias in the dataset has been accounted for in the machine learning process. Therefore, it is important to consider machine learning models as clinical decision support to alleviate the concern for reduction in revenue for healthcare organizations or rendering the clinicians less important [25].

In conclusion, our systematic review reveals the potential of machine learning models in the management of oral cancer. More importantly, resolving the issues related to the concerns highlighted in this systematic review will ensure a faster implementation of this approach in clinical practice. This would further enhance an informed clinical decision-making and offer a better diagnosis, treatment and prognostication of oral cancer.

Authors Contribution

Study concepts and study design: Alabi RO, Elmusrati M, Almangush A, Leivo I. **Studies extraction:** Alabi RO, Omar Y (would be acknowledge in acknowledgment). **Acquisition and quality control of included studies:** Alabi RO, Almangush A. **Data analysis and interpretation:** Alabi RO, Elmusrati M, Almangush A, Mäkitie AA, Pirinen M, Leivo I. **Manuscript preparation:** Alabi RO, Almangush A, Mäkitie AA, Pirinen M. **Manuscript review:** Mäkitie AA, Leivo I, Elmusrati M, Pirinen M. **Manuscript editing:** Almangush, Alabi RO. All authors approved the final manuscript for submission.

Summary points

What was already known on the topic:

- There are published studies on the application of machine learning techniques to analyse oral tongue squamous cell carcinoma (OTSCC).
- The machine model used in actual clinical practice is limited due to certain limitations and concerns.

What knowledge this study adds:

- To the best of our knowledge, this is the first study that systematically review the published studies that examined the application of machine learning techniques to analyse tongue squamous cell carcinoma (OTSCC).
- It examines the concerns and limitations to the actual implementation of machine learning-based models in clinical settings. This study also offers possible solutions to these concerns.
- Support vector machine and artificial neural network are the most widely used algorithms for oral cancer prognostication.

- Addressing these limitations as suggested in this study may ensure that the models are useful for effective oral cancer management.

Acknowledgment

The School of Technology and Innovations, University of Vaasa Scholarship Fund. The Helsinki University Hospital Research Fund.

Figure Legend

Figure 1. The flow diagram highlighting the search strategy and the search results**.**

Figure 2. Machine learning training scheme showing the concern to actual implementation.

Figure 3. The black-box concern of the machine learning models in oral cancer management

Figure 4. An explainable and trustworthy machine learning model.

References

[1] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2020, CA. Cancer J. Clin. 70 (2020) 7–30. https://doi.org/10.3322/caac.21590.

[2] S. Huang, J. Yang, S. Fong, Q. Zhao, Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges, Cancer Lett. 471 (2020) 61–71. https://doi.org/10.1016/j.canlet.2019.12.007.

[3] L. Zhu, W. Luo, M. Su, H. Wei, J. Wei, X. Zhang, C. Zou, Comparison between artificial neural network and Cox regression model in predicting the survival rate of gastric cancer patients, Biomed. Rep. 1 (2013) 757–760. https://doi.org/10.3892/br.2013.140.

[4] J. Faradmal, A.R. Soltanian, G. Roshanaei, R. Khodabakhshi, A. Kasaeian, Comparison of the Performance of Log-logistic Regression and Artificial Neural Networks for Predicting Breast Cancer Relapse, Asian Pac. J. Cancer Prev. 15 (2014) 5883–5888. https://doi.org/10.7314/APJCP.2014.15.14.5883.

[5] R.O. Alabi, M. Elmusrati, I. Sawazaki-Calone, L.P. Kowalski, C. Haglund, R.D. Coletta, A.A. Mäkitie, T. Salo, I. Leivo, A. Almangush, Machine learning application for prediction of locoregional recurrences in early oral tongue cancer: a Web-based prognostic tool, Virchows Arch. 475 (2019) 489–497. https://doi.org/10.1007/s00428-019-02642-5.

[6] C.-W. Chien, Y.-C. Lee, T. Ma, T.-S. Lee, Y.-C. Lin, W. Wang, W.-J. Lee, The application of artificial neural networks and decision tree model in predicting post-operative complication for gastric cancer patients, Hepatogastroenterology. 55 (2008) 1140–1145.

[7] M.R. Gohari, A. Biglarian, E. Bakhshi, M.A. Pourhoseingholi, Use of an artificial neural network to determine prognostic factors in colorectal cancer patients, Asian Pac. J. Cancer Prev. APJCP. 12 (2011) 1469–1472.

[8] K. Kourou, T.P. Exarchos, K.P. Exarchos, M.V. Karamouzis, D.I. Fotiadis, Machine learning applications in cancer prognosis and prediction, Comput. Struct. Biotechnol. J. 13 (2015) 8–17. https://doi.org/10.1016/j.csbj.2014.11.005.

[9] H.M. Zolbanin, D. Delen, A. Hassan Zadeh, Predicting overall survivability in comorbidity of cancers: A data mining approach, Decis. Support Syst. 74 (2015) 150–161. https://doi.org/10.1016/j.dss.2015.04.003.

[10] D. Chen, K. Xing, D. Henson, L. Sheng, A.M. Schwartz, X. Cheng, Developing Prognostic Systems of Cancer Patients by Ensemble Clustering, J. Biomed. Biotechnol. 2009 (2009) 1–7. https://doi.org/10.1155/2009/632786.

[11] C. Denkert, G. von Minckwitz, S. Darb-Esfahani, B. Lederer, B.I. Heppner, K.E. Weber, J. Budczies, J. Huober, F. Klauschen, J. Furlanetto, W.D. Schmitt, J.-U. Blohmer, T. Karn, B.M. Pfitzner, S. Kümmel, K. Engels, A. Schneeweiss, A. Hartmann, A. Noske, P.A. Fasching, C. Jackisch, M. van Mackelenbergh, P. Sinn, C. Schem, C. Hanusch, M. Untch, S. Loibl, Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy, Lancet Oncol. 19 (2018) 40–50. https://doi.org/10.1016/S1470-2045(17)30904-X.

[12] Y. Mintz, R. Brodie, Introduction to artificial intelligence in medicine, Minim. Invasive Ther. Allied Technol. 28 (2019) 73–81. https://doi.org/10.1080/13645706.2019.1575882.

[13] Z. Qian, Y. Li, Y. Wang, L. Li, R. Li, K. Wang, S. Li, K. Tang, C. Zhang, X. Fan, B. Chen, W. Li, Differentiation of glioblastoma from solitary brain metastases using radiomic machine-learning classifiers, Cancer Lett. 451 (2019) 128–135. https://doi.org/10.1016/j.canlet.2019.02.054.

[14] A. Tan, H. Huang, P. Zhang, S. Li, Network-based cancer precision medicine: A new emerging paradigm, Cancer Lett. 458 (2019) 39–45. https://doi.org/10.1016/j.canlet.2019.05.015.
[15] R.O. Alabi, M. Elmusrati, I. Sawazaki-Calone, L.P. Kowalski, C. Haglund, R.D. Coletta, A.A. Mäkitie, T. Salo, A. Almangush, I. Leivo, Comparison of supervised machine learning classification techniques in prediction of locoregional recurrences in early oral tongue cancer, Int. J. Med. Inf. (2019) 104068. https://doi.org/10.1016/j.ijmedinf.2019.104068.
[16] P.M. Speight, A.E. Elliott, J.A. Jullien, M.C. Downer, J.M. Zakzrewska, The use of artificial intelligence to identify people at risk of oral cancer and precancer, Br. Dent. J. 179 (1995) 382–387. https://doi.org/10.1038/sj.bdj.4808932.
[17] N. Sharma, H. Om, Usage of Probabilistic and General Regression Neural Network for Early Detection and Prevention of Oral Cancer, Sci. World J. 2015 (2015) 1–11. https://doi.org/10.1155/2015/234191.
[18] N. Sharma, H. Om, GMDH polynomial and RBF neural network for oral cancer classification, Netw. Model. Anal. Health Inform. Bioinforma. 4 (2015). https://doi.org/10.1007/s13721-015-0085-2.
[19] M. Aubreville, C. Knipfer, N. Oetter, C. Jaremenko, E. Rodner, J. Denzler, C. Bohr, H. Neumann, F. Stelzle, A. Maier, Automatic Classification of Cancerous Tissue in Laserendomicroscopy Images of the Oral Cavity using Deep Learning, Sci. Rep. 7 (2017). https://doi.org/10.1038/s41598-017-12320-8.
[20] R.D. Uthoff, B. Song, S. Sunny, S. Patrick, A. Suresh, T. Kolur, G. Keerthi, O. Spires, A. Anbarani, P. Wilder-Smith, M.A. Kuriakose, P. Birur, R. Liang, Point-of-care, smartphone-based, dual-modality, dual-view, oral cancer screening device with neural network classification for low-resource communities, PLOS ONE. 13 (2018) e0207493. https://doi.org/10.1371/journal.pone.0207493.
[21] M. Al-Ma'aitah, A.A. AlZubi, Enhanced Computational Model for Gravitational Search Optimized Echo State Neural Networks Based Oral Cancer Detection, J. Med. Syst. 42 (2018). https://doi.org/10.1007/s10916-018-1052-0.
[22] K. Lalithamani, A. Punitha, Detection of oral cancer using deep neural based adaptive fuzzy system in data mining techniques., Int. J. Recent Technol. Eng. 7 (2019) 397–405.
[23] K.P. Exarchos, Y. Goletsis, D.I. Fotiadis, Multiparametric Decision Support System for the Prediction of Oral Cancer Reoccurrence, IEEE Trans. Inf. Technol. Biomed. 16 (2012) 1127–1134. https://doi.org/10.1109/TITB.2011.2165076.
[24] C.-S. Cheng, P.-W. Shueng, C.-C. Chang, C.-W. Kuo, Adapting an Evidence-based Diagnostic Model for Predicting Recurrence Risk Factors of Oral Cancer, J. Univers. Comput. Sci. 24 (2018) 742–752.
[25] A.M. Bur, A. Holcomb, S. Goodwin, J. Woodroof, O. Karadaghy, Y. Shnayder, K. Kakarala, J. Brant, M. Shew, Machine learning to predict occult nodal metastasis in early oral squamous cell carcinoma, Oral Oncol. 92 (2019) 20–25. https://doi.org/10.1016/j.oraloncology.2019.03.011.
[26] M. Mermod, E. Jourdan, R. Gupta, M. Bongiovanni, G. Tolstonog, C. Simon, J. Clark, Y. Monnier, Development and validation of a multivariable prediction model for the identification of occult lymph node metastasis in oral squamous cell carcinoma, Head Neck. (2020). https://doi.org/10.1002/hed.26105.
[27] N. Sharma, H. Om, Data mining models for predicting oral cancer survivability, Netw. Model. Anal. Health Inform. Bioinforma. 2 (2013) 285–295. https://doi.org/10.1007/s13721-013-0045-7.

[28] W.-T. Tseng, W.-F. Chiang, S.-Y. Liu, J. Roan, C.-N. Lin, The Application of Data Mining Techniques to Oral Cancer Prognosis, J. Med. Syst. 39 (2015). https://doi.org/10.1007/s10916-015-0241-3.
[29] C. Lu, J.S. Lewis, W.D. Dupont, W.D. Plummer, A. Janowczyk, A. Madabhushi, An oral cavity squamous cell carcinoma quantitative histomorphometric-based image classifier of nuclear morphology can risk stratify patients for disease-specific survival, Mod. Pathol. 30 (2017) 1655–1665. https://doi.org/10.1038/modpathol.2017.98.
[30] O.A. Karadaghy, M. Shew, J. New, A.M. Bur, Development and Assessment of a Machine Learning Model to Help Predict Survival Among Patients With Oral Squamous Cell Carcinoma, JAMA Otolaryngol. Neck Surg. 145 (2019) 1115. https://doi.org/10.1001/jamaoto.2019.0981.
[31] B. Zheng, S.W. Yoon, S.S. Lam, Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms, Expert Syst. Appl. 41 (2014) 1476–1482. https://doi.org/10.1016/j.eswa.2013.08.044.
[32] C.M. Lynch, B. Abdollahi, J.D. Fuqua, A.R. de Carlo, J.A. Bartholomai, R.N. Balgemann, V.H. van Berkel, H.B. Frieboes, Prediction of lung cancer patient survival via supervised machine learning classification techniques, Int. J. Med. Inf. 108 (2017) 1–8. https://doi.org/10.1016/j.ijmedinf.2017.09.013.
[33] R. Al-Bahrani, A. Agrawal, A. Choudhary, Colon cancer survival prediction using ensemble data mining on SEER data, in: 2013 IEEE Int. Conf. Big Data, IEEE, Silicon Valley, CA, USA, 2013: pp. 9–16. https://doi.org/10.1109/BigData.2013.6691752.
[34] J.P. Anderson, J.R. Parikh, D.K. Shenfeld, V. Ivanov, C. Marks, B.W. Church, J.M. Laramie, J. Mardekian, B.A. Piper, R.J. Willke, D.A. Rublee, Reverse Engineering and Evaluation of Prediction Models for Progression to Type 2 Diabetes: An Application of Machine Learning Using Electronic Health Records, J. Diabetes Sci. Technol. 10 (2015) 6–18. https://doi.org/10.1177/1932296815620200.
[35] D. Moher, A. Liberati, J. Tetzlaff, D.G. Altman, PRISMA Group, Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement, PLoS Med. 6 (2009) e1000097. https://doi.org/10.1371/journal.pmed.1000097.
[36] W. Luo, D. Phung, T. Tran, S. Gupta, S. Rana, C. Karmakar, A. Shilton, J. Yearwood, N. Dimitrova, T.B. Ho, S. Venkatesh, M. Berk, Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View, J. Med. Internet Res. 18 (2016) e323. https://doi.org/10.2196/jmir.5870.
[37] V. Romeo, R. Cuocolo, C. Ricciardi, L. Ugga, S. Cocozza, F. Verde, A. Stanzione, V. Napolitano, D. Russo, G. Improta, A. Elefante, S. Staibano, A. Brunetti, Prediction of Tumor Grade and Nodal Status in Oropharyngeal and Oral Cavity Squamous-cell Carcinoma Using a Radiomic Approach, Anticancer Res. 40 (2020) 271–280. https://doi.org/10.21873/anticanres.13949.
[38] C.-Y. Wang, T. Tsai, H.-M. Chen, C.-T. Chen, C.-P. Chiang, PLS-ANN based classification model for oral submucous fibrosis and oral carcinogenesis, Lasers Surg. Med. 32 (2003) 318–326. https://doi.org/10.1002/lsm.10153.
[39] T. Kawazu, K. Araki, S. Kanda, Application of neural networks to the prediction of lymph node metastasis in oral cancer, Int. Congr. Ser. 1230 (2001) 1295–1296. https://doi.org/10.1016/S0531-5131(01)00258-8.
[40] S.K. Majumder, N. Ghosh, P.K. Gupta, Relevance vector machine for optical diagnosis of cancer, Lasers Surg. Med. 36 (2005) 323–333. https://doi.org/10.1002/lsm.20160.
[41] G.S. Nayak, S. Kamath, K.M. Pai, A. Sarkar, S. Ray, J. Kurien, L. D'Almeida, B.R. Krishnanand, C. Santhosh, V.B. Kartha, K.K. Mahato, Principal component analysis and artificial neural network analysis of oral tissue fluorescence spectra: Classification of

normal premalignant and malignant pathological conditions, Biopolymers. 82 (2006) 152–166. https://doi.org/10.1002/bip.20473.
[42] K.-Y. Kim, I.-H. Cha, A novel algorithm for lymph node status prediction of oral cancer before surgery, Oral Oncol. 47 (2011) 1069–1073. https://doi.org/10.1016/j.oraloncology.2011.07.017.
[43] S.-W. Chang, S. Abdul-Kareem, A.F. Merican, R.B. Zain, Oral cancer prognosis based on clinicopathologic and genomic markers using a hybrid of feature selection and machine learning methods, BMC Bioinformatics. 14 (2013). https://doi.org/10.1186/1471-2105-14-170.
[44] S.-W. Chang, A. Sameem, A.M. Amir Feisal Merican, Z. Rosnah Binti, A Hybrid Prognostic Model for Oral Cancer based on Clinicopathologic and Genomic Markers, Sains Malays. 43 (2014) 567–573.
[45] N. Sharma, H. Om, Using MLP and SVM for predicting survival rate of oral cancer patients, Netw. Model. Anal. Health Inform. Bioinforma. 3 (2014). https://doi.org/10.1007/s13721-014-0058-x.
[46] W. Shams, Z. Htike, Oral cancer prediction using gene expression profilling and machine learning, Int. J. Appl. Eng. Res. 12 (2017) 4893–4898.
[47] T. Turki, Z. Wei, Boosting support vector machines for cancer discrimination tasks, Comput. Biol. Med. 101 (2018) 236–249. https://doi.org/10.1016/j.compbiomed.2018.08.006.
[48] D.K. Das, S. Bose, A.K. Maiti, B. Mitra, G. Mukherjee, P.K. Dutta, Automatic identification of clinically relevant regions from oral tissue histological images for oral squamous cell carcinoma diagnosis, Tissue Cell. 53 (2018) 111–119. https://doi.org/10.1016/j.tice.2018.06.004.
[49] A. Nawandhar, N. Kumar, V. R, L. Yamujala, Stratified squamous epithelial biopsy image classifier using machine learning and neighborhood feature selection, Biomed. Signal Process. Control. 55 (2020) 101671. https://doi.org/10.1016/j.bspc.2019.101671.
[50] H. Yan, M. Yu, J. Xia, L. Zhu, T. Zhang, Z. Zhu, Tongue squamous cell carcinoma discrimination with Raman spectroscopy and convolutional neural networks, Vib. Spectrosc. 103 (2019) 102938. https://doi.org/10.1016/j.vibspec.2019.102938.
[51] M. Yu, H. Yan, J. Xia, L. Zhu, T. Zhang, Z. Zhu, X. Lou, G. Sun, M. Dong, Deep convolutional neural networks for tongue squamous cell carcinoma classification using Raman spectroscopy, Photodiagnosis Photodyn. Ther. 26 (2019) 430–435. https://doi.org/10.1016/j.pdpdt.2019.05.008.
[52] C.-H. Chan, T.-T. Huang, C.-Y. Chen, C.-C. Lee, M.-Y. Chan, P.-C. Chung, Texture-Map-Based Branch-Collaborative Network for Oral Cancer Detection, IEEE Trans. Biomed. Circuits Syst. 13 (2019) 766–780. https://doi.org/10.1109/TBCAS.2019.2918244.
[53] A. Zlotogorski-Hurvitz, B.Z. Dekel, D. Malonek, R. Yahalom, M. Vered, FTIR-based spectrum of salivary exosomes coupled with computational-aided discriminating analysis in the diagnosis of oral cancer, J. Cancer Res. Clin. Oncol. 145 (2019) 685–694. https://doi.org/10.1007/s00432-018-02827-6.
[54] L. Lavanya, J. Chandra, Oral cancer analysis using machine learnig techniques, Int. J. Eng. Res. Technol. 12 (2019) 596–601.
[55] X. Wang, J. Yang, C. Wei, G. Zhou, L. Wu, Q. Gao, X. He, J. Shi, Y. Mei, Y. Liu, X. Shi, F. Wu, J. Luo, Y. Guo, Q. Zhou, J. Yin, T. Hu, M. Lin, Z. Liang, H. Zhou, A personalized computational model predicts cancer risk level of oral potentially malignant disorders and its web application for promotion of non-invasive screening, J. Oral Pathol. Med. (2020). https://doi.org/10.1111/jop.12983.

[56] S. Sunny, A. Baby, B.L. James, D. Balaji, A. N. V., M.H. Rana, P. Gurpur, A. Skandarajah, M. D'Ambrosio, R.D. Ramanjinappa, S.P. Mohan, N. Raghavan, U. Kandasarma, S. N., S. Raghavan, N. Hedne, F. Koch, D.A. Fletcher, S. Selvam, M. Kollegal, P.B. N., L. Ladic, A. Suresh, H.J. Pandya, M.A. Kuriakose, A smart tele-cytology point-of-care platform for oral cancer screening, PLOS ONE. 14 (2019) e0224885. https://doi.org/10.1371/journal.pone.0224885.

[57] P.R. Jeyaraj, E.R. Samuel Nadar, Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm, J. Cancer Res. Clin. Oncol. 145 (2019) 829–837. https://doi.org/10.1007/s00432-018-02834-7.

[58] Y. Ariji, M. Fukuda, Y. Kise, M. Nozawa, Y. Yanashita, H. Fujita, A. Katsumata, E. Ariji, Contrast-enhanced computed tomography image assessment of cervical lymph node metastasis in patients with oral cancer by using a deep learning system of artificial intelligence, Oral Surg. Oral Med. Oral Pathol. Oral Radiol. 127 (2019) 458–463. https://doi.org/10.1016/j.oooo.2018.10.002.

[59] S. Xu, Y. Liu, W. Hu, C. Zhang, C. Liu, Y. Zong, S. Chen, Y. Lu, L. Yang, E.Y.K. Ng, Y. Wang, Y. Wang, An Early Diagnosis of Oral Cancer based on Three-Dimensional Convolutional Neural Networks, IEEE Access. 7 (2019) 158603–158611. https://doi.org/10.1109/ACCESS.2019.2950286.

[60] M.P. McRae, S.S. Modak, G.W. Simmons, D.A. Trochesset, A.R. Kerr, M.H. Thornhill, S.W. Redding, N. Vigneswaran, S.K. Kang, N.J. Christodoulides, C. Murdoch, S.J. Dietl, R. Markham, J.T. McDevitt, Point-of-care oral cytology tool for the screening and assessment of potentially malignant oral lesions, Cancer Cytopathol. (2020). https://doi.org/10.1002/cncy.22236.

[61] M.K. Yu, J. Ma, J. Fisher, J.F. Kreisberg, B.J. Raphael, T. Ideker, Visible Machine Learning for Biomedicine, Cell. 173 (2018) 1562–1565. https://doi.org/10.1016/j.cell.2018.05.056.

[62] B. Heinrichs, S.B. Eickhoff, Your evidence? Machine learning algorithms for medical diagnosis and prediction, Hum. Brain Mapp. (2019). https://doi.org/10.1002/hbm.24886.

[63] A. Altmann, L. Toloşi, O. Sander, T. Lengauer, Permutation importance: a corrected feature importance measure, Bioinformatics. 26 (2010) 1340–1347. https://doi.org/10.1093/bioinformatics/btq134.

[64] P.A. Keane, E.J. Topol, With an eye to AI and autonomous diagnosis, Npj Digit. Med. 1 (2018). https://doi.org/10.1038/s41746-018-0048-y.

[65] I. Kononenko, Machine learning for medical diagnosis: history, state of the art and perspective, Artif. Intell. Med. 23 (2001) 89–109. https://doi.org/10.1016/S0933-3657(01)00077-X.

[66] S. Patil, K. Habib Awan, G. Arakeri, C. Jayampath Seneviratne, N. Muddur, S. Malik, M. Ferrari, S. Rahimi, P.A. Brennan, Machine learning and its potential applications to the genomic study of head and neck cancer—A systematic review, J. Oral Pathol. Med. 48 (2019) 773–779. https://doi.org/10.1111/jop.12854.

[67] G. Levitin, Computational Intelligence in Reliability Engineering Evolutionary Techniques in Reliability Analysis and Optimization, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. http://link.springer.com/book/10.1007/978-3-540-37368-1 (accessed February 25, 2020).

[68] D. Michie, D.J. Spiegelhalter, C.C. Taylor, eds., Machine learning, neural and statistical classification, Ellis Horwood, New York, 1994.

[69] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature. 521 (2015) 436–444. https://doi.org/10.1038/nature14539.

[70] A. Shaban-Nejad, M. Michalowski, D.L. Buckeridge, Health intelligence: how artificial intelligence transforms population and personalized health, Npj Digit. Med. 1 (2018). https://doi.org/10.1038/s41746-018-0058-9.
[71] A.L. Fogel, J.C. Kvedar, Artificial intelligence powers digital medicine, Npj Digit. Med. 1 (2018). https://doi.org/10.1038/s41746-017-0012-2.
[72] D. Castelvecchi, Can we open the black box of AI., Nature. 538 (2016) 20–23.
[73] Bernease Herman, The Promise and Peril of Human Evaluation for Model Interpretability, (2017). http://interpretable.ml/ (accessed January 15, 2020).
[74] European Union, High-Level Expert Group on Artificial Intelligence: Ethics guidelines for trustworthy AI, (2019).
[75] L. Zachary, The Doctor Just Won't Accept That!, (2017). http://interpretable.ml/ (accessed January 15, 2020).
[76] K. Grace, J. Salvatier, A. Dafoe, B. Zhang, O. Evans, When Will AI Exceed Human Performance? Evidence from AI Experts, ArXiv170508807 Cs. (2018). http://arxiv.org/abs/1705.08807 (accessed January 16, 2020).
[77] N.D. Shah, E.W. Steyerberg, D.M. Kent, Big Data and Predictive Analytics: Recalibrating Expectations, JAMA. 320 (2018) 27. https://doi.org/10.1001/jama.2018.5602.

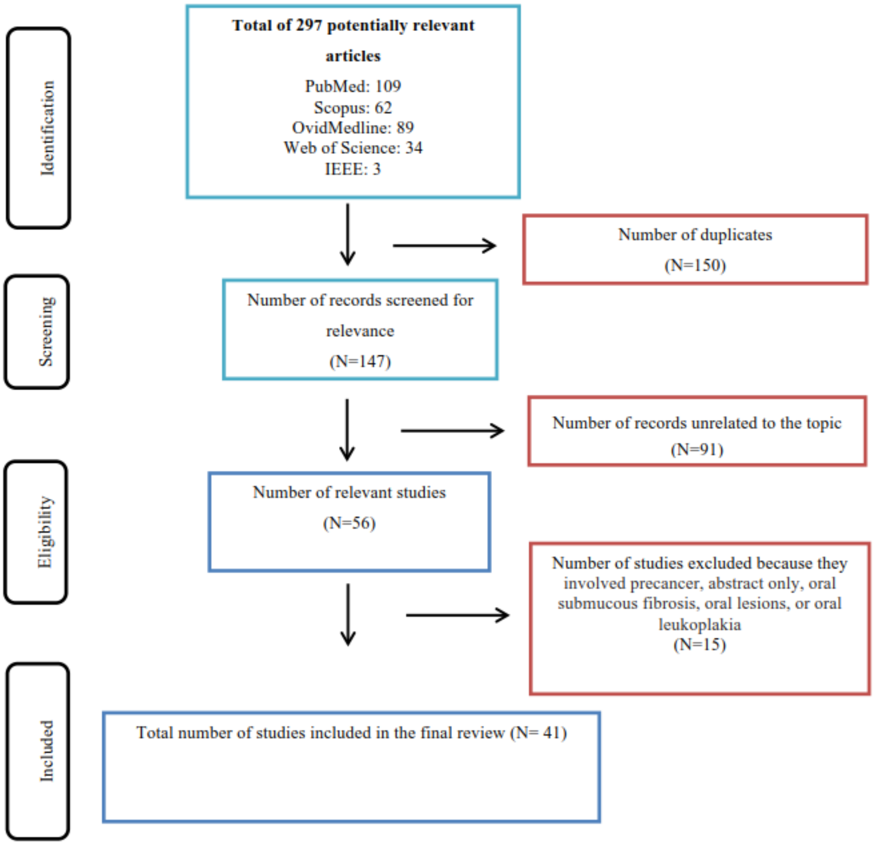

Figure 1. The flow diagram highlighting the search strategy and the search results.

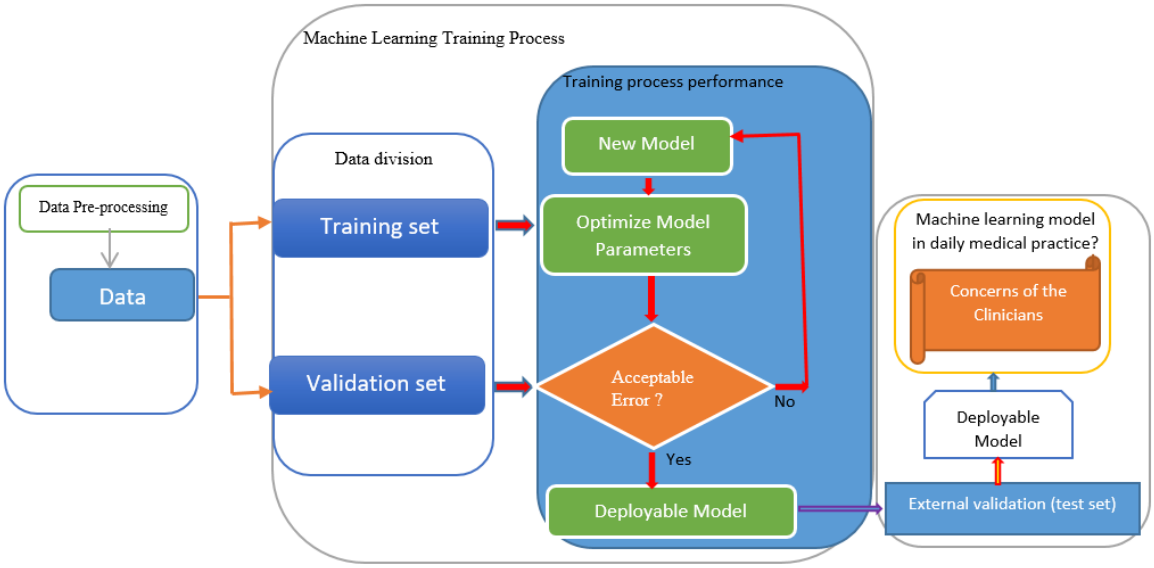

Figure 2. Machine learning training scheme showing the concern to actual implementation.

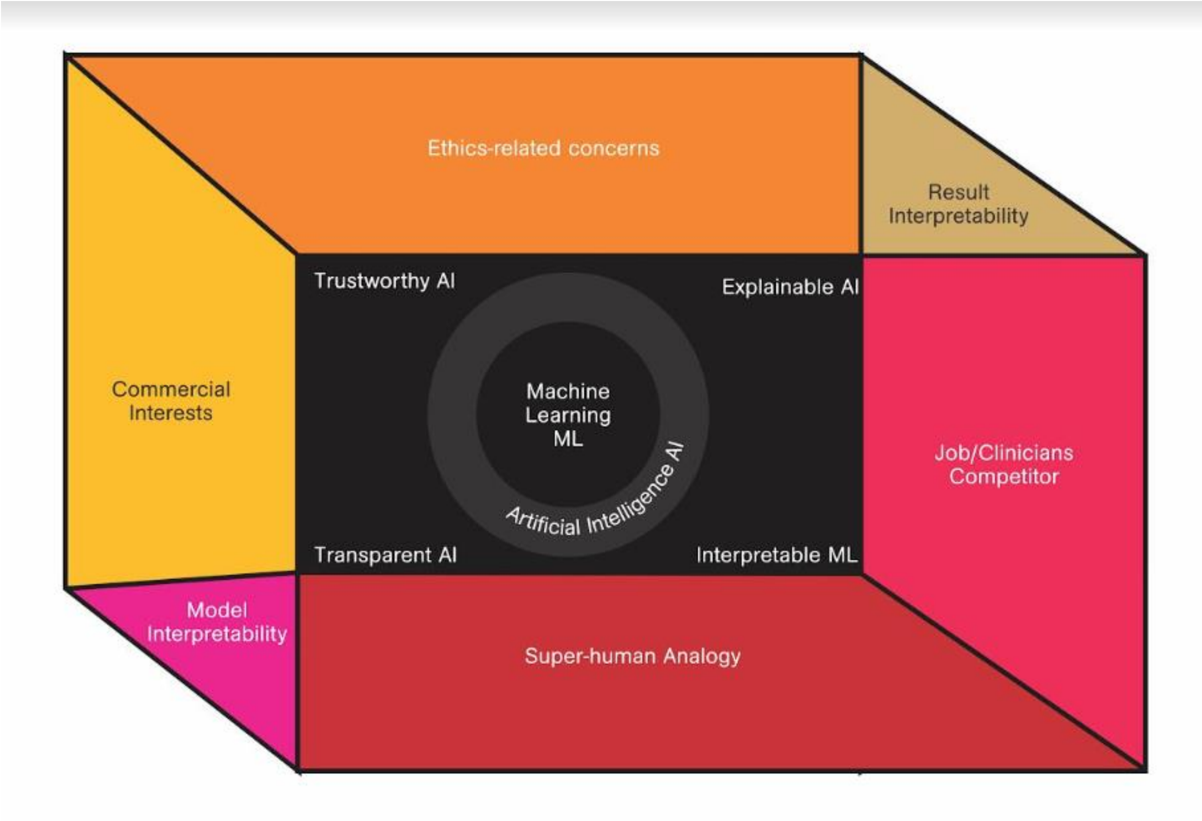

Figure 3. The black-box concern of the machine learning models in oral cancer management

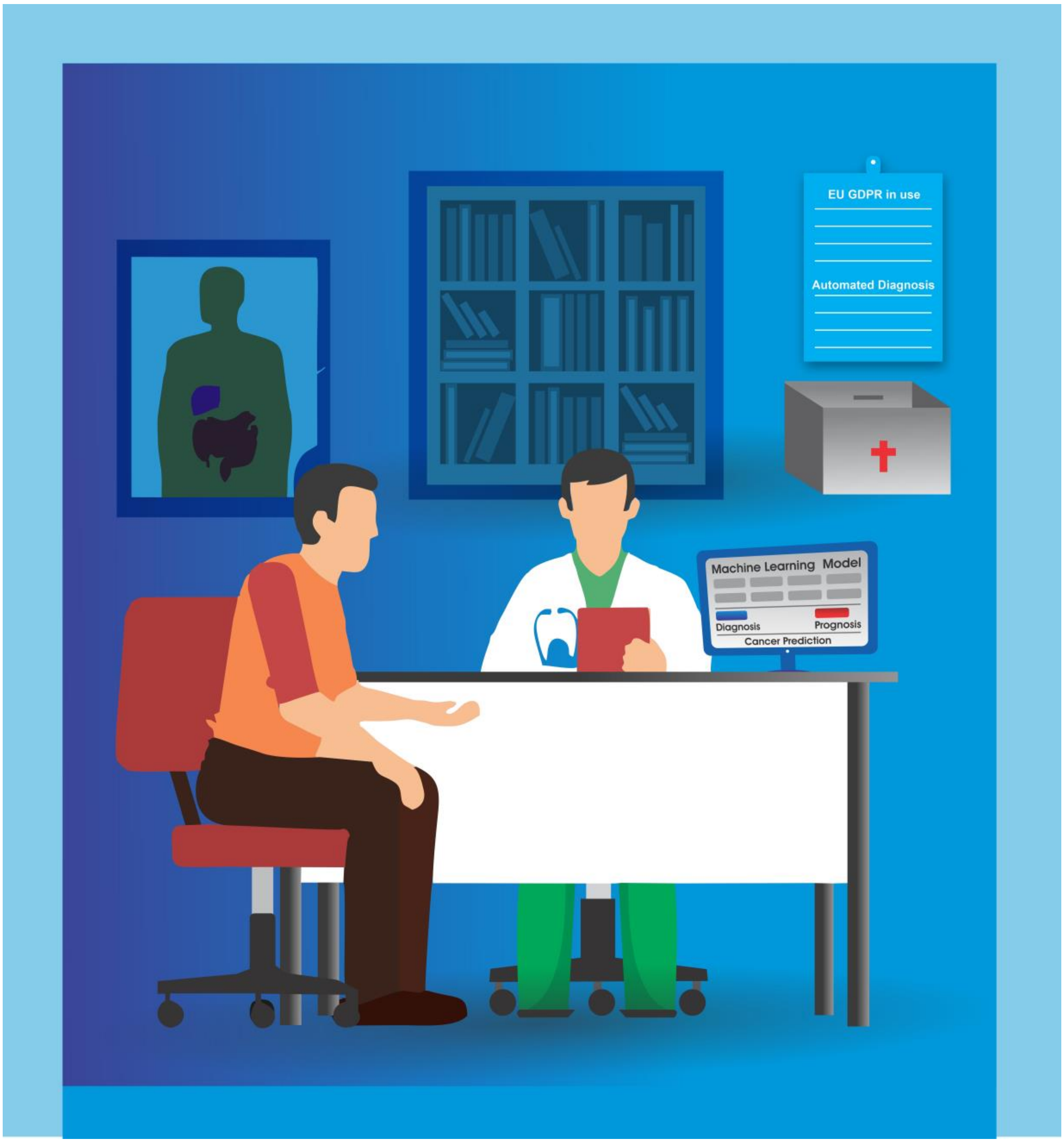

Figure 4. An explainable and trustworthy machine learning model.

Table 1. Extracts of the main findings from the included studies

Authors, year (country of authors)	Site	No of Cases	Machine Learning Methods	Use of Machine Learning in Oral cancer	Study Aim	Results	Performance metric(s)
Speight et al., 1995 (United Kingdom)	Oral cavity	2027	Neural Network	Diagnostic (data of risk habits, personal details, dental attendance).	To predict the likelihood of an individual to having a malignant or potentially malignant oral lesion.	This approach showed promising results compared with the performance of the dentist for the screening exercise.	Sensitivity: 0.80 Specificity: 0.77
Wang et al., 2003 (China)	Oral cavity*	97	Partial Least Squares and Artificial Neural Network (PLS-ANN)	Diagnostic (autofluorescence spectra data analysis).	To differentiate between premalignant and malignant tissues from benign.	The multivariate algorithm differentiated human premalignant and malignant lesions from benign lesions or normal oral mucosa.	Sensitivity: 0.81 Specificity: 0.96
Kawazu et al., 2003 (Japan)	Oral cavity	1,116	Neural Network	Diagnostic (Histopathological)	To predict lymph node metastasis in oral cancer	The prediction performance was comparable to clinical radiologists	Sensitivity: 0.80 Specificity: 0.94 Accuracy: 93.6%
Majumder et al., 2005 (India)	Oral cavity	171	Relevance Vector Machine (RVM) & Support Vector Machine (SVM)	Diagnostic (autofluorescence spectra data analysis)	To diagnose early stage oral cancer	The performance shown by the Bayesian framework of RVM was comparable to the traditional SVM.	Sensitivity: 0.91 Specificity: 0.96

Nayak et al., 2006 (India)	Oral cavity	143	Principal Component Analysis (PCA) & Artificial Neural Network (ANN)	Diagnostic (autoflourescence spectra data analysis).	To classify images into normal, premalignant, and malignant.	The performance of ANN was better than PCA.	Sensitivity: 0.96 Specificity: 1.00
Kim & Cha, 2011 (Korea)	Oral cavity	90	Principal Component Analysis (PCA)	Prognostic (Clinical and genomic)	To predict lymph node status before surgery	The model performed better when the clinical and genomic parameters were combined.	Sensitivity: 0.70 Specificity: 0.88 Accuracy: 84.0%
Exarchos et al., 2012 (Greece)	Oral cavity	41	Bayesian Networks (BN), Artificial Neural Network (ANN), Support Vector Machine (SVM), Decision Tree (DT) & Random Forest (RF)	Prognostic (Clinical, image and genomic).	To predict oral cancer reoccurrence.	The multiparametric approach presented successfully predicted oral cancer reoccurrence.	Accuracy: 100%
Sharma and Om, 2013 (India)	Oral cavity	1024	Single Tree (ST), Decision Tree Forest (DTF), Tree Boost (TB) model	Prognostic (clinicopathologic)	To predict the survival rate in cancer patients.	The three examined algorithms showed similar results and performances.	Sensitivity: 1.00 Specificity: 1.00
Chang *et al.*, 2013 (Malaysia)	Oral cavity	31	Adaptive Neuro Fuzzy Inference System (ANFIS), Artificial Neural Network (ANN), Support Vector Machine (SVM), Logistic Regression (LR)	Prognostic (Clinicopathologic and genomic)	Oral cancer prognosis using the hybrid of feature selection and several machine learning methods. [*Continuation of previous studies*]	Prognosis is more accurate with the combination of clinicopathologic and genomic markers.	Accuracy: 93.8%
Chang et al., 2014 (Malaysia)	Oral cavity	31	ReliefF-Genetic Algorithm, Feature Selection, Adaptive Neuro Fuzzy Inference System (ANFIS	Prognostic (Clinicopathologic and genomic)	To apply the hybrid of feature selection (Relief-GA) & machine learning technique (ANFIS) in prognosis of oral cancer.	The prognoses was more accurate in group 2 (clinicopa	Accuracy: 93.8%

						thologic and genomic) than group 1 (clinicopathologic markers only)	
Sharma and Om, 2014 (India)	Oral cavity	1024	Support Vector Machine (SVM) & Multi-layer Perceptron (MLP)	Prognostic (Clinicopathologic)	To predict survivability of oral cancer patients.	The performance metrics showed by SVM outperforms the multi-layer perceptron.	Sensitivity: 0.73 Specificity: 0.73 Accuracy: 73.6%
Tseng et al., 2015 (Taiwan)	Oral cavity	673	Decision Tree (DT), Artificial Neural Network (ANN), Logistic Regression (LR), & K-means	Prognostic (Clinicopathologic)	To predict 5-year survival rate and recurrence. Clustering of patients were conducted.	Decision tree and neural network showed superior to traditional method.	Accuracy: 98.4%
Sharma and Om, 2015 (India)	Oral cavity	1025	Probabilistic and General Neural Network (PNN/GRNN), Linear Regression (LR), Decision Tree (DT), Tree Boost (TB), Multi-layer perceptron (MLP), Convolutional Neural Network (CNN)	Diagnostic (Clinicopathologic)	To detect oral cancer.	The model predicted cancer stages and survivability	Sensitivity: 0.92 Specificity: 0.79 Accuracy: 80.0%
Sharma & Om, 2015 (India)	Oral cavity	1025	Group method if data handling (GMDH) polynomial neural network & Radial basis neural network (RBNN)	Diagnostic (Clinicopathologic)	To diagnose new cases of oral cancer.	The two variant of NN showed competitive results in differentiating patients with or without oral cancer.	Sensitivity: 0.77 Specificity: 0.61 Accuracy: 67.8%
Shams & Htike, 2017 (Malaysia)	Oral cavity	86	Support Vector Machine (SVM), Deep Neural Network (DNN),	Prognostic (Gene expression data).	To predict the risks of oral cancer in oral premalignant	The DNN technique performed better	Sensitivity:0.98 Specificity: 0.94 Accuracy: 96%

			Regularized Least Squares (RLS) & Multi-layer perceptron (MLP)		lesion (OPL) patients.	than others.	
Aubreville *et al.*, 2017 (Germany)	Oral cavity	7,894	Deep learning technologies on Confocal Laser Endomicroscopy (CLE) images of oral squamous cell carcinoma (OSCC)	Diagnostic (image analysis)	Detection of oral cancer based on images.	A CNN-based image recognition was successfully applied on confocal laser endomicroscopy images of OSCC.	Sensitivity: 0.86 Specificity: 0.90 Accuracy: 88.3% AUC: 0.96
Lu *et al.*, 2017 (China & USA)	Oral cavity	115	Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Support Vector Machine (SVM), Random Forest (RF)	Prognostic (Clinicopathologic + image analysis).	To predict the disease-specific survival.	The study properly associated local nuclear morphologic heterogeneity with long term outcomes.	AUC: 0.72
Uthoff *et al.*, 2018 (USA & India)	Oral cavity	170	Convolutional Neural Network (CNN)	Diagnostic (image analysis)	Early detection of precancerous and cancerous lesions	A low-cost, smartphone-based image system for oral screening was developed	Sensitivity: 0.85 Specificity: 0.88
Al-Ma'aitah & AlZubi, 2018 (Saudi Arabia)	Oral cavity	-	Gravitational Search Optimized Echo State Neural Networks (GSOESNN, Support Vector Machine (SVM), Multi-layer perceptron (MLP), & Neural Network	Diagnostic (image analysis)	Detection of oral cancer	The optimized neural network examined in this study identified oral cancer than other machine learning methods.	Accuracy: 99.2%.

Turki & Wei, 2018 (Saudi Arabia & USA)	Oral cavity*	86	Boosted Support Vector Machine (BSVM)	Prognostic (gene expression data)	Identification of oral cancer	The boosting versions of the examined algorithms outperformed the baseline algorithms.	MAUC: 0.849.
Cheng *et al.*, 2018 (Taiwan)	Oral cavity	1,429	K-Nearest Neighbor (KNN), K-shortest paths (K-STAR), Randomizable Filtered Classifier (RFC), & Random Tree (RT)	Diagnostic (Clinicopathological data)	To predict recurrence	Important risk factors for recurrence were identified. Also, KSTAR algorithm showed the best performance	Specificity: 0.75 Accuracy: 77.0%
Das *et al.*, 2018 (India)	Oral cavity	126	Deep Convolution Neural Network (DCNN)	Diagnostic (image analysis)	Automatic identification of relevant regions for OSCC diagnosis	Keratin pearls region were identified with significant accuracy.	Accuracy: 96.9%
Nawandhar *et al.,* 2019 (India)	Oral cavity	676	Decision Tree (DT), Quadratic Support Vector Machine (QSVM), Cubic SVM (Cu-SVM), Neighborhood Component Analysis (NCA), Random-Subspaces Linear Discriminant Analysis (RS-LDA) & Stratified Squamous Epithelium – Biopsy Image Classifier (SSC–BIC)	Prognostic (Image analysis)	To develop an automatic OSCC image classifier	H&E stained microscopic images were classified as either normal, well, moderately, or poorly differentiated	Accuracy: 95.6%

Yan et al., 2019 (China)	Tongue Squamous Cell Carcinoma (TSCC)	24	Convolutional Neural Networks (CNN)	Diagnostic (Raman Spectroscopy)	To discriminate the border of tongue squamous cell carcinoma from non-tumorous tissue.	The extracted features combined to produce significan t accuracy for tongue squamou s cell carcinom a discrimin ations	Sensitivity: 0.99 Specificity: 0.95 Accuracy: 97.2%
Yu et al., 2019 (China)	Oral Tongue Squamous Cell Carcinoma (OTSCC)	36	Deep Convolutional Neural Networks (DCNN), Principle Component Analysis (PCA), Support Vector Machine (SVM), & Linear Discriminant Analysis (LDA)	Diagnostic (Raman spectral data)	To discriminate OTSCC from non-tumorous tissue	DCNN showed better result than the state-of-the-art methods	Sensitivity: 0.99 Specificity: 0.94 Accuracy: 96.9%
Chan *et al.*, 2019 (Taiwan)	Oral cavity	80	Deep Convolutional Neural Networks (DCNN)	Diagnostic (auto-fluorescence data analysis)	To detect oral cancer	The feature extracted by Gabor filter provide more useful informati on for cancer detection	Sensitivity: 0.93 Specificity: 0.94
Bur *et al.*, 2019 (USA)	Oral cavity	782	Decision Forest (DF), Gradient Boosting (GB)	Prognostic (clinicopathologic)	Predict occult nodal metastasis	The DF and GB performe d better at predictin g occult nodal metastasi s than DOI model.	Specificity: 0.57 Accuracy: 63.4%
Zlotogorski-Hurvitz *et al.*, 2019 (Israel)	Oral cavity	34	Principal Component Analysis – Linear Discriminant Analysis (PCA-LDA), Support Vector Machine (SVM)	Prognostic (saliva samples)	To differentiate between the spectra of oral cancer and healthy individuals.	The mid-infrared (IR) spectra of oral cancer patients was different	Specificity: 89% Accuracy: 95%

						from healthy individuals. The PCA-LDA outperformed other examined techniques.	
Alabi *et al., 2019* (Finland &Brazil)	Oral Tongue Squamous Cell Carcinoma (OTSCC)	254	Support Vector Machine (SVM), Naive Bayes (NB), Boosted Decision Tree (BDT), Decision Forest (DF), & Permutation Feature Importance (PFI)	Prognostic (clinicopathologic)	To predict locoregional recurrence	The BDT produced the highest accuracy. Also, the examined algorithms performed better than the depth of invasion model.	Sensitivity: 0.79 Specificity: 0.83 Accuracy: 81%
Lalithamani *et al.,* 2019 (India)	Oral cavity	-	Deep Neural Based Adaptive Fuzzy System (DNAFS)	Diagnostic (demographics and histopathologic)	To identify oral cancer patients	The novel classifier uses fuzzy logic and DNN for oral cancer identification and detection	Accuracy: 96.3%
Lavanya & Chandra, 2019 (India)	Oral cavity	-	Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Multi-layer perceptron (MLP), Logistic Regression (LR)	Prognostic (Pathological data)	To classify oral cancer into stages	The ML predicted different stages in oral cancer	Accuracy: 90.6%
Wang *et al.,* 2019 (China)	Oral cavity	266	Random Forest (RF)	Prognostic (personal details, smoking & drinking status, lesion conditions, & histological grade)	Predict cancer risk of oral potentially malignant disorders.	The personalized model performed better than the baseline & clinical expert	Sensitivity: 0.82 Specificity: 0.91
Alabi *et al.,* 2019 (Finland & Brazil)	Oral tongue squamous cell	311	Artificial Neural Network (ANN)	Prognostic (Clinicopathological data)	Prediction of locoregional recurrences	The accuracy of the	Sensitivity: 0.71 Specificity: 0.98 Accuracy: 88.2%

	carcinoma (OTSCC)					neural network was significantly higher.	
Karadaghy *et al.*, 2019 (USA)	Oral cavity	33,065	Decision Forest (DF)	Prognostic (Clinicopathological, social and demographic data)	Prediction of 5-year overall survival of OSCC patients	Combining clinicopathological, social and demographics produced better model than TNM-based model.	Accuracy: 71%
Sunny et al., 2019 (India, Germany & America)	Oral cavity	100	Artificial Neural Network (ANN)	Diagnostic (image) & prognostic (clinicopathologic)	To develop a risk stratification model using ANN. Also to enable tele-cytology-based point of care diagnosis (detection of OPML).	The ANN showed higher accuracy.	Specificity: 0.90 Accuracy: 86%
Jeyaraj & Samuel Nadar, 2019 (India)	Oral cavity	100	Convolution Neural Network (CNN)	Diagnostic (image analysis)	To use CNN for the detection of cancerous tumor with benign and cancerous tumor with normal tissue.	The regression-based partitioned CNN performs better than other traditional medical image classification technique examined.	Sensitivity: 0.94 Specificity: 0.91 Accuracy: 91.4 %
Ariji et al., 2019 (Japan)	Oral cavity	45	Convolution Neural Network (CNN)	Diagnostic (image analysis)	To evaluate the performance of CNN for the diagnosis of lymph node metastasis.	The CNN yielded performance that is similar to pathologists.	Sensitivity: 0.75 Specificity: 0.81 Accuracy: 78.2%.
Xu et al., 2019 (China)	Oral cavity	~ 7000	Three-Dimensional Convolutional Neural Networks (3DCNN)	Diagnostic (image analysis)	To differentiate between benign and malignant oral cancers	The 3DCNN variant gave a better performa	Accuracy: 75.4%

						nce than the 2DCNN in differentiating between benign and malignant.	
Romeo *et al.*, 2020 (Italy)	Oral cavity	40	Naïve Bayes (NB), Bagging of NB, K-Nearest Neighbors (KNN), J48, boosting J48	Prognostic (Image analysis)	Prediction of tumor grade and nodal status in patients with OCSCC & oropharyngeal.	Most accurate subset of features to predict tumor grade and nodal status were identified.	Accuracy: 92.9%
McRae et al., 2020 (USA)	Oral cavity	999	K-Nearest Neighbors (KNN)	Diagnostic (histopathologic and brush cytologic parameters)	To detect potential malignant oral lesions (PMOL).	This approach represent a practical solution for quick PMOL assessment.	Accuracy: 99.3%
Mermod et al., 2020 (Switzerland & Australia)	Oral cavity	56 (112 external validation)	Random Forest (RF), linear Support Vector Machine (SVM), LASSO regularized logistic regression, C5.0 decision trees	Prognostic (demographic, histopathologic, immunohistochemical)	To predict occult lymph node metastases (OLNM)	The examined algorithm offered a clinical management strategies to identify patients that would benefit from neck dissection	Sensitivity: 0.8 Specificity: 0.9 Accuracy: 90%

Table 2. Quality measurement guidelines [Adapted from Luo et al., 2016] [36]

Article sections	Parameters	Explanation
Title	▪ Title (Nature of Study)	The study clearly showed that it focused on either diagnostic or prognosis model, or both.
Abstract	• Abstract (Structured summary of the study)	It contains the background, objectives, data sources, performance metrics and conclusion. The data sources and no of data is preferred but can also be optional in the abstract.
Introduction	▪ Rationale ▪ Objectives	Describes the goals of the study. It properly introduced the reader to the study. A brief introduction that reviews the current practice and prediction performance of existing models. Also, identify how the newly proposed model may benefit the clinical practices.
Methods	▪ Describe the available data/describe the setting ▪ Define the problem (diagnostic/prognostic) ▪ Data preparation ▪ Build the model	Describe the data source, size of data sample, year/duration of the available data. The nature of the data (retrospective/prospective), input and target variables definition, cost of prediction errors, performance metrics definition, and the explanation of the success criteria. Data inclusion and exclusion criteria, data processing methods, missing values and how it was handled. Finally, explain how the model was built. (Explaining the nature of data and the external validation are desirable but not mandatory)
Results	• The performance of the model using the external validation dataset	This reports the final model and its performance. It is recommended to compare the performance of the model with other known models, clinical standards or statistical methods. Reporting the confidence intervals is optional but desirable. Similarly, it is highly recommended to validate the model externally. If not possible, internal validation becomes important.
Discussion	▪ Discuss the clinical implications ▪ Discuss the limitations	Discuss the significance of the findings and possible limitations (potential pitfalls) of the study or the model to be specific. Mentioning the financial implications, that is, the amount of money that can be saved using this model is optional.
Conclusion	○ Discuss the overall usage of the model in the clinical arena.	Report the unexpected signs of the model such as collinearity, overfitting, underfitting. Most importantly, evaluates if the objective of the studies was fulfilled.

Table 3. Quality scores of the included studies based on the guidelines provided Luo et al., 2016 [36, guidelines modified]

Studies	Tittle	Abstract	Rationale	Objectives	Setting Description	Problem Definition	Data Preparation	Build Model	Report Performance	Clinical Implications	Limitations	Scores (%)
Speight et al., 1995	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🔵	90.0%
Wang et al., 2003	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🔵	🟠	90.0%
Majumder et al., 2005	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🔵	🟠	90.0%
Nayak et al., 2006	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	100.0%
Exarchos et al., 2012	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🔵	90.0%
Sharma & Ohm, 2013	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🔵	🔵	81.8%
Chang et al., 2013	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🔵	🔵	81.8%
Chang et al., 2014	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🔵	90.0%
Tseng et al., 2015	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	100.0%
Sharma & Ohm, 2015	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	100.0%
Sharma & Om, 2015	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🔵	🔵	81.8%
Shams & Htike, 2017	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🔵	🔵	81.8%
Aubreville et al., 2017	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	100.0%
Lu et al., 2017	🔵	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	90.0%
Uthoff et al., 2018	🔵	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🔵	81.8%
Al-Ma'aitah & Alzubi, 2018	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🔵	🔵	81.8%
Turki & Wei, 2018	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🔵	🔵	81.8%
Cheng et al., 2018	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🔵	90.0%
Das et al., 2018	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🔵	90.0%
Nawandhar et al., 2019	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🔵	90.0%
Yu et al., 2019	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🔵	90.0%
Chan et al., 2019	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🔵	🔵	81.8%
Bur et al., 2019	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	100.0%
Zlotogorski-Hurvitz et al., 2019	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🔵	90.0%
Alabi et al., 2019	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	100.0%
Lalithamani et al., 2019	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🔵	🟠	100.0%
Lavanya & Chandra, 2019	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🔵	90.0%
Wang et al., 2019	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	90.0%
Alabi et al., 2019	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	100.0%
Karadaghy et al., 2019	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	100.0%
Sunny et al., 2019	🔵	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	90.0%
Jeyaraj & Samuel Nadar., 2019	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🔵	90.0%
Ariji et al., 2019	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	100.0%
Xu et al., 2019	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🔵	90.0%
Romeo et al., 2020	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	100.0%
McRae et al., 2020	🔵	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	90.0%
Mermod et al., 2020	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	🟠	100.0%