Comparison of nomogram with machine learning techniques for prediction of overall survival in patients with tongue cancer

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

Title: Comparison of nomogram with machine learning techniques for prediction of overall survival in patients with tongue cancer

Year: 2021

Version: Final draft (post print, aam, accepted manuscript)

Please cite the original version:

Alabi, R. O., Mäkitie, A. A., Pirinen, M., Elmusrati, M., Leivo, I., & Almangush, A., (2021). Comparison of nomogram with machine learning techniques for prediction of overall survival in patients with tongue cancer. International journal of medical informatics 145. https://doi.org/10.1016/j.ijmedinf.2020.104313

Comparison of nomogram with machine learning techniques for prediction of overall survival in patients with tongue cancer

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

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

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

[c] 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.

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

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

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

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

Abstract

Background: The prediction of overall survival in tongue cancer is important for planning of personalized care and patient counselling. **Objectives:** This study compares the performance of a nomogram with a machine learning model to predict overall survival in tongue cancer. The nomogram and machine learning model were built using a large data set from the Surveillance, Epidemiology, and End Results (SEER) program database. The comparison is necessary to provide the clinicians with a comprehensive, practical, and most accurate assistive system to predict overall survival of this patient population. **Methods:** The data set used included the records of 7596 tongue cancer patients. The considered machine learning algorithms were logistic regression, support vector machine, Bayes point machine, boosted decision tree, decision forest, and decision jungle. These algorithms were mainly evaluated in terms of the areas under the receiver-operating characteristic (ROC) curve (AUC) and accuracy values. The performance of the algorithm that produced the best result was compared with a nomogram to predict overall survival in tongue cancer patients. **Results:** The boosted decision-tree algorithm outperformed other algorithms. When compared with a nomogram using external validation data, the boosted decision tree produced an accuracy of 88.7% while the nomogram showed an accuracy of 60.4%. In addition, it was found that age of patient, T stage, radiotherapy, and the surgical resection were the most prominent features with significant influence on the machine learning model's performance to predict overall survival**. Conclusion:** The machine learning model provides more personalized and reliable prognostic information of tongue cancer than the nomogram. However, the level of transparency offered by the nomogram in estimating patients' outcomes seems more confident and strengthened the principle of shared decision making between the patient and clinician. Therefore, a combination of a nomogram – machine learning (NomoML) predictive model may help to improve care, provides information to patients, and facilitates the clinicians in making tongue cancer management-related decisions.

KEYWORDS: Machine learning; Nomogram; tongue cancer; Predict; overall survival

1.   Introduction

The most common epithelial neoplasm affecting the oral cavity is the oral cavity squamous cell carcinoma (OCSCC) [1]. Tongue squamous cell carcinoma (TSCC) accounts for the most common cases in oral cavity cancers [2–6]. It is characterized by an aggressive clinical behavior [7] such as rapid local invasion and early lymph node metastasis [8]. This aggressive behavior leads to a have high rate of recurrence and mortality [9]. Despite the advancement in cancer diagnostic and management approaches in recent years, the 5-year relative overall survival (OS) for patients treated with curative intent was 61% in a recent study [10] and 63% in another report [11].

The prediction of tongue cancer survival outcomes is of utmost interest to both clinicians and patients. This is because determining cancer outcomes may crucially contribute to personalized treatment planning and even to avoiding unnecessary therapies [12]. Also, it provides a useful insight into effective management decision-making and may guide the selection of a protocol treatment approach. Of note, predicting TSCC survival is challenging due to different patient-related factors, tumor characteristics, and available treatment modalities. The American Joint Committee on Cancer (AJCC) Tumor-Nodal-Metastasis (TNM) staging system has shown to be an objective and accurate tool for predicting the prognosis for an entire population of cancer patients. Thus, it was widely used for planning of treatment strategies for TSCC patients [8,13]. However, for an individual patient, it is ineffective for predicting outcome due to its inability to consider other tumor- and patient-related risk factors [14,15]. To this end, a tool that considers these factors together to accurately predict patients' outcomes would be pertinent [8].

Nomogram is defined as a pictorial representation of a complex mathematical formula that uses certain variables such as demographics, clinical, or treatment variables to graphically depict a statistical prognostic model [16,17]. This graphical representation of the prognostic

model can be used for the prognostication of clinical events such as recurrence, disease-specific survival or overall survival for a given patient [17]. Nomograms have been used in predicting survival in breast cancer [18], gastric cancer [19], and head and neck cancer [8,20,21]. Similarly, machine learning techniques have been touted for effective prediction of outcomes. These include for instance, predicting locoregional recurrence [22,23], occult node metastasis [24,25], and survival rates [26–29].

In this study, we aim to compare the performance of a nomogram with machine learning techniques in predicting the overall survival of tongue cancer patients. The survival time in months of tongue cancer patients was considered as the time from the beginning of treatment until the last follow-up or death [30]. The examined machine learning algorithms were logistic regression, support vector machine (SVM), naive Bayes (NB), neural network, boosted decision tree, decision forest, and decision jungle algorithms. This comparison is pertinent as it is aimed at providing the clinicians with a comprehensive, practical, and most accurate assistive system to predict overall survival for patients. Additionally, this system will assist the clinicians to provide a more personalized and precise therapeutic decision. This study is based on multi-population data obtained from the National Cancer Institute (NCI) through the Surveillance, Epidemiology, and End Results (SEER) Program of the National Institutes of Health (NIH).

2. Material and Methods

2.1. National Cancer Institute Database

The study data were obtained from the National Cancer Institute (NCI) through the Surveillance, Epidemiology, and End Results (SEER) Program of the National Institutes of Health (NIH). It is one of the largest cancer database that is available publicly [31]. It gives non-identifiable information on cancer statistics of the United States population. These

important characteristics of this database make it a database of choice, thus facilitating large-scale outcome analysis research. The ethical permission to use the SEER database was approved with the user identification numbers of 10455-Nov2108 and 11522-Nov2019, respectively.

2.2. Selection of patient attributes

The SEER database was chosen as it was considered as a high-quality database of different cancer patients [8]. The SEER program of the National Cancer Institute was searched for Nov 2015 submission [1973 – 2013] (Figure 1). These years (1973 – 2013) under consideration were also selected because the nomogram to be used for comparison was built using the same date range.

The inclusion criteria included that the patients were diagnosed with histologically confirmed (positive) tongue cancer. Additionally, the patient must have known basic data such as gender and age at diagnosis. Therefore, the included known clinical and pathologic characteristics were age at diagnosis, race, marital status, grade, TNM status according to AJCC 7th edition, treatment (surgery, and radiotherapy) [8]. The survival period (in months) and overall survival status of the patients were also extracted. All patients whose diagnostic information were unknown were excluded. A total of 7649 cases were found eligible to be included in this study (Table 1). The data extraction process is shown in Figure 1. The explanation of the included variables and categorization is shown in Table 2.

2.3. Separate external validation cases: Out of 7649 cases, we reserved the last 53. These cases were not used in the machine learning training and testing phase. It was reserved to externally validate the model that showed the best performance metrics in terms of the accuracy. The external validation cases, who had been labelled as dead, had died within 5 years from the first treatment. Similarly, the individuals who were labelled as alive were alive at least

5 years from the first treatment. It is important to externally validate the model to address the possible concern about the generalizability of the model.

2.4. Nomogram

We used the nomogram constructed in another previously published study for evaluating the 5- and 8-year overall survival in tongue squamous cell carcinoma (Figure 2-3) [8]. It was chosen because it considered overall survival as a distinct event in its construction. Additionally, it was well-validated (internal and external validation) and calibrated [8].

2.5. Machine learning training process

 Microsoft Azure Machine Learning Studio (Azure ML 2019) was used in this study [32]. The input parameters were age at diagnosis, race, sex, marital status at diagnosis, tumor grade, AJCC TNM staging system, survival time, treatment (radiotherapy, surgical resection). The output variable was overall survival (alive or dead). The survival months ranged from 1 to 47 months. Firstly, the extracted data were checked to ensure that these were properly preprocessed. In addition, all the variables were converted to numeric to reduce possible spelling errors and omissions in each variable (Table 2). Also, potential class imbalance in the target variable was handled by up-sampling using synthetic minority oversampling technique (SMOTE) [33]. This approach offers a reasonable approach to handling potential imbalance than simply duplicating existing cases.

The data set was divided into two sets of training and validation using a 5-fold cross-validation method in the ratio 80% training and 20% validation {80:20} percentage splitting sets [34] (Figure 4). Using cross-validation, hyperparameters were fine-tuned to maximize the area under receiving operating characteristics curve (AUC) or concordant partial AUC especially when imbalanced dataset was used in the training [35] for each of the examined algorithms. Each of the algorithms of interest was then configured and used for the whole training data set [36,37]. After training, the algorithms were evaluated for the performance

metrics of interest (Figure 4, Table 3). The performance of classification algorithms was compared mainly in terms of accuracy, AUC (internal validation), F1 score, and likelihood ratios as represented in Table 3. The algorithm that showed the best AUC values was used for external validation and comparison with the nomogram. The data used for external validation were not used during the training phase. The result obtained for validating this model externally was considered as the true performance of the algorithm (Table 4). Also, it addressed possible concerns relating to the generalization of the algorithm.

2.6. **Comparison of the performance of the machine learning with a nomogram:** The nomogram and the machine learning algorithm that showed the best accuracy were compared using the external validation data (Section 2.3). The machine learning algorithm was compared with a nomogram built with surgical treatment (Figure 2). The result of this comparison is presented in Table 4. Likewise, the algorithm was compared with a nomogram built with radiotherapy (Figure 3). The result of this comparison is given in Table 4. The overall performance of these two predictive tools in terms of accuracy, sensitivity, specificity, and F1 score is shown in Table 4. The comparison was necessary to ensure that the predictive tool used in medicine is convenient, accurate, and explainable (enables clinicians to understand why the algorithm produced certain result). This was corroborated by the study of Holzinger et al., where human and machine explanations were compared using system causability scale (SCS) to allow for explainable AI [38].

3. **Results**

3.1. **Data Description**

The study cohort included 7596 patients with tongue squamous cell carcinoma; 5322 male and 2274 female in a male-to-female ratio of 2.3:1. The mean age at diagnosis was 62.3 (SD $\pm$ 12.7: range 12-102) and the median age was 62.0 years. In terms of the ethnicity, 6597 (86.8%) were

from the white origin, 516 (6.8%) were black, and 483 (6.4%) were from other origins including American Indian/AK Native, Asian/Pacific Islander. Considering marital status, 4430 (58.3%) were married while 3166 (41.7%) were considered unmarried (single, divorced, widowed, and separated) at the time of diagnosis. For 1215 (16.0%) out of the 7596 patients the grade was well-differentiated, for 3768 (49.6%) moderately differentiated, 2543 (33.5%) poorly differentiated, and 70 (0.9%) undifferentiated.

Similarly, the distribution according to the AJCC TNM staging scheme, the tumor diameter showed that 2942 (38.7%) patients had stage T1, 2492 (32.8%) stage T2, 1159 (15.3%) stage T3, 920 (12.1%) stage T4a, and 83 (1.1%) stage T4b. Also, 3485 (45.9%) had N0, 1220 (16.1%) had N1, 327 (4.3%) N2a, 1498 (19.7%) N2b, 880 (11.6%) N2c, 186 (2.4%) N3; 7425 M0, and 171 M1. The histopathologic characteristics are briefly summarized in Table 2. In terms of the treatment, 4654 (61.3%) had surgery while 2942 (38.7) had no surgery. Adjuvant radiotherapy was administered to 4489 (59.1%) patients. The follow-up time ranged from 0 to 47 months (Mean 18.3; SD $\pm$ 13.3). The number of patients who were alive at last follow-up was 5743 (75.6%).

For the testing series (n = 53) that was not included in the construction of the machine learning model, the mean age at diagnosis was 61.1 years (SD $\pm$ 11.1; range from 31 to 85 years with 44 (83.0%) male and 9 (17.0%) female. Additionally, 38 (71.7%) were married and 15 (28.3%) unmarried. For histological grade, 29 (54.7%) were well-differentiated, 10 (18.9%) moderately-differentiated, 9 (17.0%) poorly-differentiated, and 5 (9.4%) undifferentiated grade. In terms of treatment, 37 (69.8%) patients had radiotherapy while 22 (41.5%) had surgery performed. The mean follow-up time was 4.2 months (SD $\pm$ 5.2; range 0-23 months) and 47 (88.6%) patients were alive at the end of follow-up. The detailed characteristics of the external validation data are given in Table 1.

3.2. Performance metrics for the algorithms

The performance metrics of the algorithms are presented in Table 3. The average specificity of the examined algorithms was 0.89. Similarly, the average sensitivity was 0.76. In terms of the accuracy and the area under receiving operating characteristics curve (AUC), the boosted decision tree outperformed all other algorithms.

3.3. Evaluating the input variables for importance

The permutation feature importance of the input variables showed that the AJCC T stage, radiotherapy, age and surgical status are the most prominent features that had significant influence on the model's performance to predict the overall survival in tongue cancer patients.

3.4. Comparison of the performance of the nomogram with machine learning algorithm

The nomogram (with surgical treatment) showed 66.0% accuracy, likewise, the nomogram (with radiotherapy) produced an accuracy of 60.4% when tested with the external validation data. The machine learning algorithm (boosted decision tree) showed an accuracy of 88.7% when tested with external validation data. All the examined methods showed 100% sensitivity (Table 4). Considering the specificity and F1 score, the nomogram (with surgical treatment) showed 0.62 and 0.40, machine learning model gave 0.87 and 0.66, and nomogram (with radiation) produced 0.55 and 0.36 (Table 4).

5. Discussion

In this study, a nomogram and several machine learning algorithms were utilized and compared in the prediction of overall survival in patients with tongue cancer. These machine learning algorithms used were logistic regression, support vector machine, naive Bayes, neural network

(NN), boosted decision tree, decision forest, and decision jungle. The algorithm that produced the best accuracy was compared with a nomogram. This comparison was based on a separate cohort that was not used in the training or testing phase. This comparison is necessary to ensure that the best model is selected for the specific management of patients with tongue cancer.

Several studies have examined the significance of machine learning (shallow and deep learning) techniques for oral cancer prognostication. For example, Tseng et al. developed a machine learning model for survival risk stratification of patients with advanced OSCC using clinicopathologic and genetic data [39]. This approach was corroborated by Karadaghy et al., where social, demographic, and clinicopathologic features were used to develop a machine learning-based model to predict 5-year overall survival of OSCC patients [29]. These studies concluded that a machine learning-based approach augments the clinicians' ability to properly estimate the survival risk of OSCC patients. Thus, effective and efficient treatment plan – intensifying or deintensifying the regimen, can be mapped out to improve the quality of care and survival of OSCC patients [39].

Besides survival risk estimate with machine learning techniques, Alabi et al., and Bur et al., have published promising results regarding predicting clinical outcome of a progressive disease such as locoregional tumor recurrence and/or distant metastases [22–24,40]. This technique has also been found to be better than such conventional methods as tumor depth of invasion (DOI), neutrophil-to-lymphocyte ratio (NLR), or tumor budding in predicting clinical outcomes [24,41]. Also, deep learning techniques have shown to be a promising noninvasive approach in early diagnosis [42], assessment of cervical lymph node metastasis [43], and discriminating between well-differentiated and poorly differentiated OSCC [44].

In this study, the boosted decision tree outperformed other algorithms and the nomogram. It uses the gradient boosting approach to create an ensemble of classification trees needed to stratify the patients in terms of their overall survival of tongue cancer. Each of the

tree created is dependent on the prior trees. The algorithm learns by fitting the errors of the tree that proceeded it. Consequently, the second tree fits the errors of the first tree, the third tree fits the errors in the second trees, the sequence of error fitting continues in that order until the final tree. Predictions are therefore based on the entire ensemble of trees [23,36,37]. With a reasonable amount of data used in the study, the boosted decision tree algorithm was able to minimize errors due to the large coverage of the relationship between the data to improve the accuracy.

Tumor stage (T stage), radiotherapy, age of patient, and surgical resection were the input variables that had significant importance on the machine learning model's ability to predict overall survival in tongue cancer patients. For the stage of the disease at diagnosis, it has been reported that it is strongly correlated with prognosis [45]. The survival of patients with stage I (T1N0) of the disease exceeds 80% while stage III-IV (T3-T4) reduces below 40% [46,47]. Interestingly, most of the oral cancer patients are usually found to be at stage III or IV at the time of diagnosis [48,49]. This further corroborated the importance of T stage on the predictive model. Similarly, the age of the patient at the time of diagnosis was found to play an important role in the model's predictive ability. This result was emphasized in other studies that reported that the survival of oral cancer patients steadily decreases with age of the patient [30,50]. As the cohort contained largely early-stage tongue cancer, it is no surprise that the treatment options had a significant impact on the predictive performance of the model. This is because the treatment of choice for early-stage tongue cancer can either be surgery, radiotherapy or combination of both [45].

Traditionally, the clinicians' judgments have formed the foundation for estimating the risk of patients, counseling and decision making. Therefore, the experience of clinicians plays a significant role in accurate risk estimation and decision making. This approach poses a great risk of bias and the predicted outcomes of the patients may be highly subjective [12,51,52].

11

The nomogram has been used to predict survival in various head and neck cancers [20,21,53–55]. Its performance was reported to provide superior disease-related risk estimations for patients [56]. Likewise, machine learning models have shown encouraging risk estimation for patients [22–24,29]. Therefore, the introduction of nomogram and machine learning models have been touted to providing the clinicians with a decision-making assistive tool that gives more accurate predictions of patients' outcome. When these two approaches were compared as presented in this study, the machine learning model outperformed the nomogram in predicting the overall survival in tongue cancer patients. To the best of our knowledge, this is the first study that compares the performance of a nomogram with machine learning for tongue cancer.

The machine learning model showed that it was able to identify and understand the hard-to-discern relationships between the input variables. The predictive accuracy exhibited by this model is particularly well-suited to medical applications for personalized and predictive medicine [57]. The boosted decision-tree algorithm was able to build formidable overall survival classification trees in a step-wise manner, where the error in each step is measured and corrected in the next step to produce a model with improved predictive accuracy. Despite the better predictive performance of the machine learning model over the nomogram, the fact that the nomogram offers an appealing, transparent means of estimating the risk of patients without the use of the internet or computer is worthy of consideration for clinical decision-making.

Of note, the transparency offered by the nomogram addresses the concerns that the results from machine learning models are not easily interpretable. With this level of transparency in calculating the patients' outcomes, it is obvious that the patient will be more confident in the recommended treatment approach. More importantly, the principle of shared decision-making between the patient and clinician can be strengthened. Therefore, a combination of nomogram – machine learning (NomoML) approach may offer a more transparent approach for individualized assessment and add to the planning of the most

appropriate adjuvant treatment for tongue cancer patients. The level of transparency in calculating the patients' outcomes in addition to the significant accuracy offered by the machine learning model is poised to give confidence to the patient for the recommended treatment approach.

In addition to the accuracy and transparency that the proposed NomoML seeks to offer, it is also poised to allow for explainable AI. However, as this proposed tool seeks to combine human and automatic (autonomous) machine learning approaches, the need to examine the causability (property of a person that measures the quality of explanations [58]) and explainability (property of a system that measures why an algorithm/system came up with certain result [38,58]) of this tool becomes imperative. An example of a viable tool to measure the quality of these explanations is the systematic causability scale (SCS) proposed by Holzinger et al. [38]. This tool (SCS) combines causability and explainability to reach the level of explainable medicine [59]. Therefore, for future study, it would be important to examine our proposed diagnostic tool for SCS evaluation. Undoubtedly, concerns about human-AI relationship and the extent to which AI-based model can or should support clinical decisions is growing. However, it is important to properly understand causability and explainability prior to addressing the former concerns [59].

In this study, there were certain limitations to be considered. Both the nomogram and machine learning were developed using retrospective cohorts; it remains important to validate these with a prospective cohort for the comparison to be a representation of the performance of these tools. Also, there may be a possibility of bias in the data set as a significant number of the patients were alive at the end of follow-up. In addition, information about some variables such as perineural invasion that have been reported to have significant influence on the overall survival are not available. Therefore, it would be worthwhile to further calibrate the nomogram to include these variables and to compare it with deep machine learning technologies.

13

In conclusion, concerted efforts should be made towards the construction of a more accurate nomogram and machine learning models. This includes use of larger data sets, inclusion of novel biomarkers, improved data collection, calibration, and validation methods. With an improved NomoML, accurate estimation of the likelihood of recurrences, tongue-specific and overall survival of tongue cancer can be greatly improved and management-related decisions for tongue cancer patients can be enhanced.

Figure Legend

Figure 1. Flowchart for data extraction for the Surveillance, Epidemiology, and End Results data selection.

Figure 2. Nomogram for predicting 5- and 8-year overall survival with surgical treatment.

Figure 3. Nomogram for predicting 5- and 8-year overall survival with radiation treatment

Figure 4. The flowchart for the machine learning process and comparison with nomogram.

Table Legend

Table 1. Clinical and tumor characteristics.

Table 2. The explanation of the included variables and categorization.

Table 3. The performance metrics of the compared machine learning algorithms.

Table 4. The performance metrics of the comparison between the nomogram and machine learning model.

Highlights

What we already know on the topic:

- Nomogram can be used to estimate the overall survival in cancer patients.

- Machine learning techniques have been used to predict the overall survival in tongue cancer patients.

What knowledge this study adds:

- To the best of our knowledge, this is the first study that compared the performance of a nomogram with machine learning techniques to estimate overall survival in tongue cancer patients.

- The machine learning model outperformed the nomogram in estimating patients' outcome.

- We proposed a combination of a nomogram – machine learning (NomoML) predictive model to improve care for tongue cancer patients.

- Improved decision-making by the clinician and improved the overall quality of care of the patients.

REFERENCES

[1] Wallace ML, Neville BW. Squamous cell carcinoma of the gingiva with an atypical appearance. J Periodontol 1996;67:1245–8. https://doi.org/10.1902/jop.1996.67.11.1245.

[2] Rodrigues VC, Moss SM, Tuomainen H. Oral cancer in the UK: to screen or not to screen. Oral Oncol 1998;34:454–65. https://doi.org/10.1016/s1368-8375(98)00052-9.

[3] Moore SR, Johnson NW, Pierce AM, Wilson DF. The epidemiology of mouth cancer: a review of global incidence. Oral Dis 2000;6:65–74. https://doi.org/10.1111/j.1601-0825.2000.tb00104.x.

[4] Marur S, Forastiere AA. Head and Neck Squamous Cell Carcinoma: Update on Epidemiology, Diagnosis, and Treatment. Mayo Clin Proc 2016;91:386–96. https://doi.org/10.1016/j.mayocp.2015.12.017.

[5] Pires FR, Ramos AB, Oliveira JBC de, Tavares AS, Luz PSR da, Santos TCRB dos. Oral squamous cell carcinoma: clinicopathological features from 346 cases from a single oral pathology service during an 8-year period. J Appl Oral Sci 2013;21:460–7. https://doi.org/10.1590/1679-775720130317.

[6] Li R, Koch WM, Fakhry C, Gourin CG. Distinct epidemiologic characteristics of oral tongue cancer patients. Otolaryngol Head Neck Surg 2013;148:792–6. https://doi.org/10.1177/0194599813477992.

[7] Bello IO, Soini Y, Salo T. Prognostic evaluation of oral tongue cancer: means, markers and perspectives (I). Oral Oncol 2010;46:630–5. https://doi.org/10.1016/j.oraloncology.2010.06.006.

[8] Li Y, Zhao Z, Liu X, Ju J, Chai J, Ni Q, et al. Nomograms to estimate long-term overall survival and tongue cancer-specific survival of patients with tongue squamous cell carcinoma. Cancer Medicine 2017;6:1002–13. https://doi.org/10.1002/cam4.1021.

[9] Almangush A, Bello IO, Coletta RD, Mäkitie AA, Mäkinen LK, Kauppila JH, et al. For early-stage oral tongue cancer, depth of invasion and worst pattern of invasion are the strongest pathological predictors for locoregional recurrence and mortality. Virchows Archiv 2015;467:39–46. https://doi.org/10.1007/s00428-015-1758-z.

[10] Mroueh R, Haapaniemi A, Grénman R, Laranne J, Pukkila M, Almangush A, et al. Improved outcomes with oral tongue squamous cell carcinoma in Finland: Oral tongue carcinoma in Finland. Head & Neck 2017;39:1306–12. https://doi.org/10.1002/hed.24744.

[11] van Dijk BAC, Brands MT, Geurts SME, Merkx MAW, Roodenburg JLN. Trends in oral cavity cancer incidence, mortality, survival and treatment in the Netherlands. Int J Cancer 2016;139:574–83. https://doi.org/10.1002/ijc.30107.

[12] Kudo Y. Predicting cancer outcome: Artificial intelligence vs. pathologists. Oral Diseases 2019;25:643–5. https://doi.org/10.1111/odi.12954.

[13] American Joint Committee on Cancer. AJCC Cancer Staging Manual. New York, NY: Springer New York; 2002. https://doi.org/10.1007/978-1-4757-3656-4.

[14] Sobin LH. TNM: Evolution and relation to other prognostic factors. Seminars in Surgical Oncology 2003;21:3–7. https://doi.org/10.1002/ssu.10014.

[15] Patel SG, Lydiatt WM. Staging of head and neck cancers: is it time to change the balance between the ideal and the practical? J Surg Oncol 2008;97:653–7. https://doi.org/10.1002/jso.21021.

[16] Grimes DA. The nomogram epidemic: resurgence of a medical relic. Ann Intern Med 2008;149:273–5. https://doi.org/10.7326/0003-4819-149-4-200808190-00010.

[17] Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. Lancet Oncol 2015;16:e173-180. https://doi.org/10.1016/S1470-2045(14)71116-7.
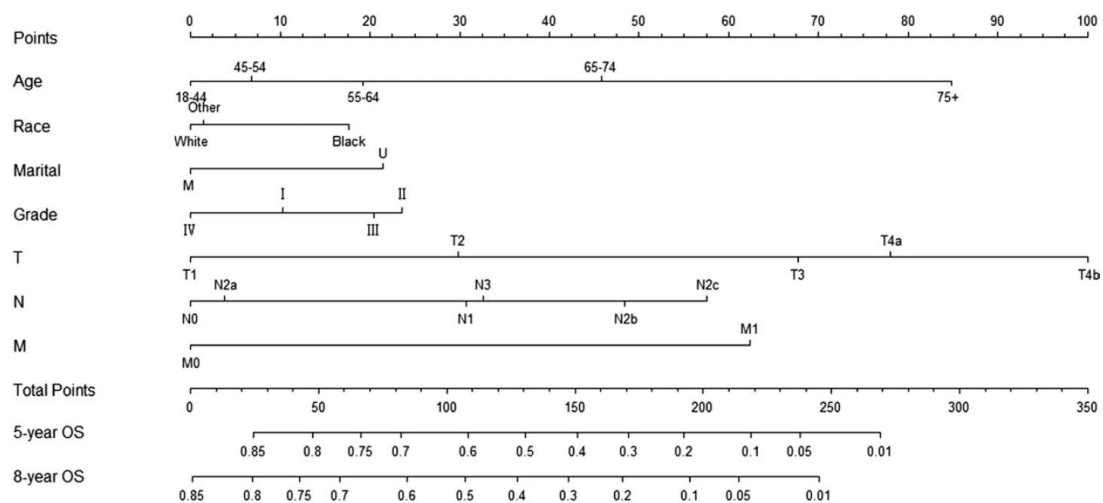
[18]   Sun W, Jiang Y-Z, Liu Y-R, Ma D, Shao Z-M. Nomograms to estimate long-term overall survival and breast cancer-specific survival of patients with luminal breast cancer. Oncotarget 2016;7:20496–506. https://doi.org/10.18632/oncotarget.7975.

[19]   Liu J, Geng Q, Liu Z, Chen S, Guo J, Kong P, et al. Development and external validation of a prognostic nomogram for gastric cancer using the national cancer registry. Oncotarget 2016;7:35853–64. https://doi.org/10.18632/oncotarget.8221.

[20]   Gross ND, Patel SG, Carvalho AL, Chu P-Y, Kowalski LP, Boyle JO, et al. Nomogram for deciding adjuvant treatment after surgery for oral cavity squamous cell carcinoma. Head Neck 2008;30:1352–60. https://doi.org/10.1002/hed.20879.

[21]   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;120:214–21. https://doi.org/10.1002/cncr.28407.

[22]   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 Archiv 2019;475:489–97. https://doi.org/10.1007/s00428-019-02642-5.

[23]   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. International Journal of Medical Informatics 2019:104068. https://doi.org/10.1016/j.ijmedinf.2019.104068.

[24]   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 Oncology 2019;92:20–5. https://doi.org/10.1016/j.oraloncology.2019.03.011.

[25]   Mermod M, Jourdan E, 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. https://doi.org/10.1002/hed.26105.

[26]   Sharma N, Om H. Data mining models for predicting oral cancer survivability. Network Modeling Analysis in Health Informatics and Bioinformatics 2013;2:285–95. https://doi.org/10.1007/s13721-013-0045-7.

[27]   Tseng W-T, Chiang W-F, Liu S-Y, Roan J, Lin C-N. The Application of Data Mining Techniques to Oral Cancer Prognosis. Journal of Medical Systems 2015;39. https://doi.org/10.1007/s10916-015-0241-3.

[28]   Lu C, Lewis JS, Dupont WD, Plummer WD, Janowczyk A, Madabhushi A. An oral cavity squamous cell carcinoma quantitative histomorphometric-based image classifier of nuclear morphology can risk stratify patients for disease-specific survival. Modern Pathology 2017;30:1655–65. https://doi.org/10.1038/modpathol.2017.98.

[29]   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 Otolaryngology–Head & Neck Surgery 2019;145:1115. https://doi.org/10.1001/jamaoto.2019.0981.

[30]   Chen S-W, Zhang Q, Guo Z-M, Chen W-K, Liu W-W, Chen Y-F, et al. Trends in clinical features and survival of oral cavity cancer: fifty years of experience with 3,362 consecutive cases from a single institution. Cancer Management and Research 2018;Volume 10:4523–35. https://doi.org/10.2147/CMAR.S171251.

[31]   SEER P. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) Research Data 1973–2009, National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, released April 2012, based on the November 2011 submission. 2012.

[32]   Microsoft Azure Machine Learning Studio. Azure Machine Learning Studio: In Documentation. 2018.

[33]   Blagus R, Lusa L. SMOTE for high-dimensional class-imbalanced data. BMC Bioinformatics 2013;14:106. https://doi.org/10.1186/1471-2105-14-106.

[34]   Tapak L, Shirmohammadi-Khorram N, Amini P, Alafchi B, Hamidi O, Poorolajal J. Prediction of survival and metastasis in breast cancer patients using machine learning classifiers. Clinical Epidemiology and Global Health 2018. https://doi.org/10.1016/j.cegh.2018.10.003.

[35]   Carrington AM, Fieguth PW, Qazi H, Holzinger A, Chen HH, Mayr F, et al. A new concordant partial AUC and partial c statistic for imbalanced data in the evaluation of machine learning algorithms. BMC Medical Informatics and Decision Making 2020;20. https://doi.org/10.1186/s12911-019-1014-6.

[36]   Microsoft Azure Machine Learning Studio. Azure Machine Learning Studio: In Documentation 2018.

[37]   Barga R, Fontama V, Tok W-H. Predictive analytics with Microsoft Azure Machine Learning. Second edition. Berkeley, CA: Apress; 2015.

[38]   Holzinger A, Carrington A, Müller H. Measuring the Quality of Explanations: The System Causability Scale (SCS): Comparing Human and Machine Explanations. KI - Künstliche Intelligenz 2020;34:193–8. https://doi.org/10.1007/s13218-020-00636-z.

[39]   Tseng Y-J, Wang H-Y, Lin T-W, Lu J-J, Hsieh C-H, Liao C-T. Development of a Machine Learning Model for Survival Risk Stratification of Patients With Advanced Oral Cancer. JAMA Network Open 2020;3:e2011768. https://doi.org/10.1001/jamanetworkopen.2020.11768.

[40]   Chu CS, Lee NP, Adeoye J, Thomson P, Choi S. Machine learning and treatment outcome prediction for oral cancer. Journal of Oral Pathology & Medicine 2020. https://doi.org/10.1111/jop.13089.

[41]   Shan J, Jiang R, Chen X, Zhong Y, Zhang W, Xie L, et al. Machine Learning Predicts Lymph Node Metastasis in Early-Stage Oral Tongue Squamous Cell Carcinoma. Journal of Oral and Maxillofacial Surgery 2020. https://doi.org/10.1016/j.joms.2020.06.015.

[42]   Jeyaraj PR, Samuel Nadar ER. Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm. Journal of Cancer Research and Clinical Oncology 2019;145:829–37. https://doi.org/10.1007/s00432-018-02834-7.

[43]   Ariji Y, Fukuda M, Kise Y, Nozawa M, Yanashita Y, Fujita H, et al. 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 Surgery, Oral Medicine, Oral Pathology and Oral Radiology 2019;127:458–63. https://doi.org/10.1016/j.oooo.2018.10.002.

[44]   Ren J, Qi M, Yuan Y, Duan S, Tao X. Machine Learning–Based MRI Texture Analysis to Predict the Histologic Grade of Oral Squamous Cell Carcinoma. American Journal of Roentgenology 2020:1–7. https://doi.org/10.2214/AJR.19.22593.

[45]   Arrangoiz R, Cordera F, Caba D, Moreno E, Luque de Leon E, Munoz M. Oral Tongue Cancer: Literature Review and Current Management. Cancer Reports and Reviews 2018;2. https://doi.org/10.15761/CRR.1000153.

[46]   Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. CA Cancer J Clin 2014;64:9–29. https://doi.org/10.3322/caac.21208.

[47]   Harrison LB, Sessions RB, Kies MS, editors. Head and neck cancer: a multidisciplinary approach. Fourth edition. Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2014.
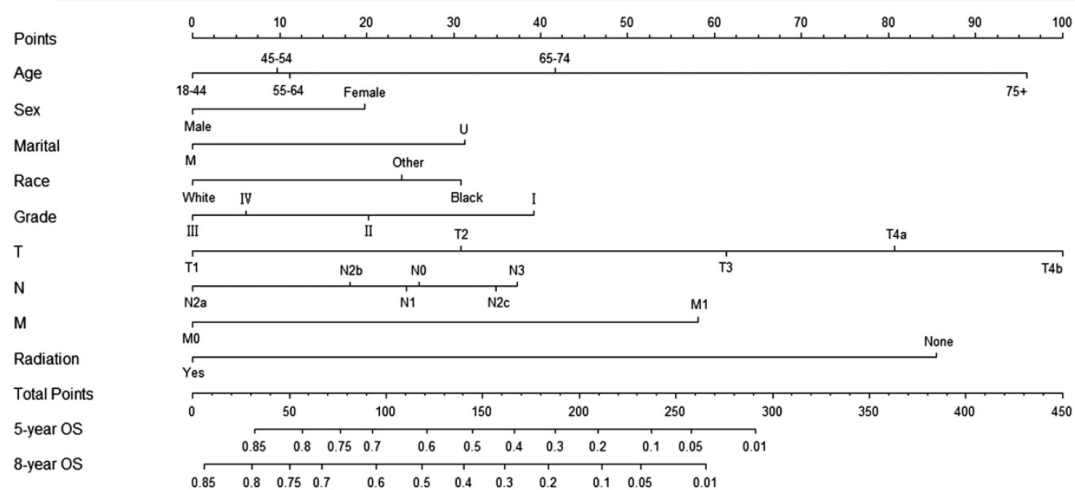
[48]   Edge SB, American Joint Committee on Cancer, American Cancer Society, editors. AJCC cancer staging handbook: from the AJCC cancer staging manual. 7th ed. New York: Springer; 2010.

[49]   Wallner PE, Hanks GE, Kramer S, McLean CJ. Patterns of Care Study. Analysis of outcome survey data-anterior two-thirds of tongue and floor of mouth. Am J Clin Oncol 1986;9:50–7. https://doi.org/10.1097/00000421-198602000-00013.

[50]   Listl S, Jansen L, Stenzinger A, Freier K, Emrich K, Holleczek B, et al. Survival of Patients with Oral Cavity Cancer in Germany. PLoS ONE 2013;8:e53415. https://doi.org/10.1371/journal.pone.0053415.

[51]   Elstein AS. Heuristics and biases: selected errors in clinical reasoning. Acad Med 1999;74:791–4. https://doi.org/10.1097/00001888-199907000-00012.

[52]   Vlaev I, Chater N. Game relativity: how context influences strategic decision making. J Exp Psychol Learn Mem Cogn 2006;32:131–49. https://doi.org/10.1037/0278-7393.32.1.131.

[53]   Ganly I, Amit M, Kou L, Palmer FL, Migliacci J, Katabi N, et al. Nomograms for predicting survival and recurrence in patients with adenoid cystic carcinoma. An international collaborative study. Eur J Cancer 2015;51:2768–76. https://doi.org/10.1016/j.ejca.2015.09.004.

[54]   Ali S, Palmer FL, Yu C, DiLorenzo M, Shah JP, Kattan MW, et al. Postoperative nomograms predictive of survival after surgical management of malignant tumors of the major salivary glands. Ann Surg Oncol 2014;21:637–42. https://doi.org/10.1245/s10434-013-3321-y.

[55]   Cho J-K, Lee G-J, Yi K-I, Cho K-S, Choi N, Kim JS, et al. Development and external validation of nomograms predictive of response to radiation therapy and overall survival in nasopharyngeal cancer patients. Eur J Cancer 2015;51:1303–11. https://doi.org/10.1016/j.ejca.2015.04.003.

[56]   Shariat SF, Karakiewicz PI, Suardi N, Kattan MW. Comparison of nomograms with other methods for predicting outcomes in prostate cancer: a critical analysis of the literature. Clin Cancer Res 2008;14:4400–7. https://doi.org/10.1158/1078-0432.CCR-07-4713.

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

[58]   Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. WIREs Data Mining and Knowledge Discovery 2019;9. https://doi.org/10.1002/widm.1312.

[59]   Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. WIREs Data Mining and Knowledge Discovery 2019;9. https://doi.org/10.1002/widm.1312.
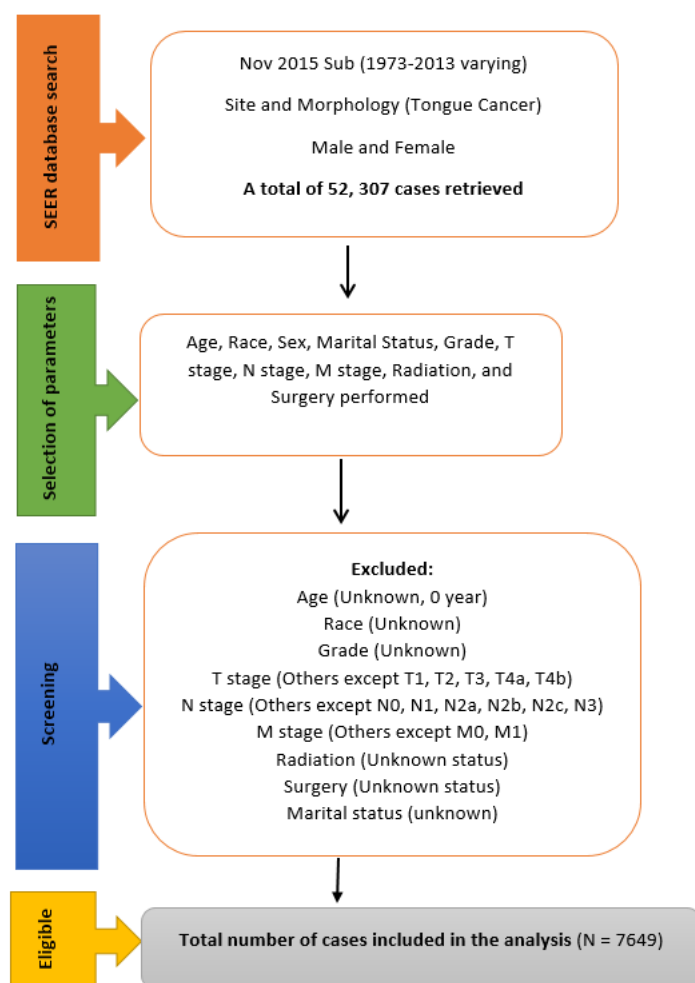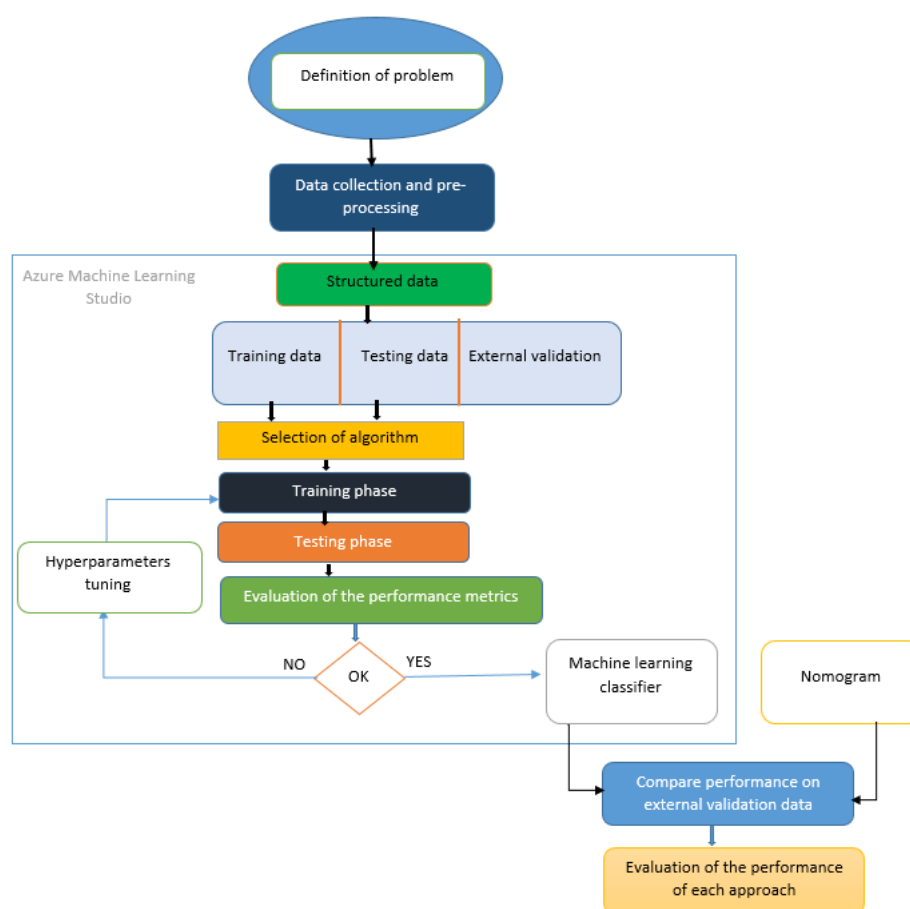
Figure 1. Flowchart for data extraction for the Surveillance, Epidemiology, and End Results data selection.

Figure 2. Nomogram for predicting 5- and 8-year overall survival with surgical treatment.

Figure 3. Nomogram for predicting 5- and 8-year overall survival with radiation treatment

Figure 4. The flowchart for the machine learning process and comparison with nomogram.

Table 1. Baseline demographic and tumor characteristics of patients in SEER database.

Variables	Overall survival, N = 7596 Training and testing cohort	Overall survival, N = 53 External validation cohort
Age at diagnosis (years)		
1 – 18	5 (0.1%)	0 (0.0%)
19 – 44	515 (6.8%)	2 (3.8%)
45 – 54	1412 (18.6%)	14 (26.4%)
55 – 64	2497 (32.9%)	18 (34.0%)
65 – 74	1877 (24.7%)	13 (24.5%)
75+	1290 (16.9%)	6 (11.3%)
Ethnic origin		
White	6597 (86.8%)	44 (83.0%)
Black	516 (6.8%)	9 (17.0%)
Other*	483 (6.4%)	
Sex		
Male	5322 (70.0%)	38 (71.7%)
Female	2274 (30.0%)	15 (28.3%)
Marital status		
Married	4430 (58.0%)	29 (54.7%)
Unmarried	3166 (42.0%)	24 (15.3%)
Grade		
Grade I	1215 (16.0%)	8 (15.1%)
Grade II	3768 (49.6%)	29 (54.7%)
Grade III	2543 (33.5%)	15 (28.3%)
Grade IV	70 (0.9%)	1 (1.9%)
T stage (2010+)		
T1	2942 (38.7%)	19 (35.8%)
T2	2492 (32.8%)	16 (30.2%)
T3	1159 (15.3%)	6 (11.3%)
T4a	920 (12.1%)	11 (20.8%)
T4b	83 (1.1%)	1 (1.9%)
N Stage (2010+)		
N0	3485 (45.9%)	14 (26.4%)
N1	1220 (16.1%)	10 (18.9%)
N2a	327 (4.3%)	5 (9.4%)
N2b	1498 (19.7%)	14 (26.4%)
N2c	880 (11.6%)	9 (17.0%)
N3	186 (2.4%)	1 (1.9%)
M stage (2010+)		
M0	7425 (97.7%)	50 (94.3%)
M1	171 (2.3%)	3 (5.7%)
Surgery performed		
Yes	4654 (61.3%)	22 (41.5%)
None	2942 (38.7%)	31 (58.5%)
Radiotherapy		
Yes	4489 (59.1%)	37 (69.8%)
None	3107 (40.9%)	16 (30.2%)
Overall survival status		
Alive	5743 (75.6%)	47 (88.7%)
Dead	1853 (24.4%)	6 (11.3%)

*Other including American Indian (native), Asian/Pacific Islander

Table 2. Selected SEER attributes, description, and categorization used in machine learning training.

Attribute	Description	Categorization for machine learning training	Type
Age	Age at time of diagnosis	No categorization	Discrete
Race/Ethnicity	This describes the ethnicity of the patient	0 = White; 1 = Black; 2= Others (American Indian /AK Native, Asian pacific	Numeric
Sex	Biological sex	0 = Male; 1 = Female	Numeric
Marital Status	The marital status of the patient at diagnosis of OTSCC.	0 = Married; 1 = Single (never married, Unmarried or domestic partner); 2 = Divorced (separated); 3 = Widowed; 4 = Separated.	Numeric
Grade	The differentiation of the cancer cell.	1 = Grade 1 (Well differentiated), 2 = Grade 2 (Moderately differentiated), 3 = Grade 3 (poorly differentiated), 4 =(Undifferentiated)	Numeric
Derived AJCC T, 7th edition (2010+) stage	T1: The tumor is ≤ 2cm or less in greatest dimension. T2: The tumor is > 2cm & ≤ 4cm. T3: Tumor is > 4cm. T4a: Moderately advanced local disease. T4b: Significantly advanced local disease.	T1 = 1; T2 =2; T3 = 3; T4a = 4; T4b = 5	Numeric
Derived AJCC N, 7th edition (2010+) stage	N0; No regional lymph node metastasis N1: Regional lymph node metastasis (single node). N2a: Cancer has spread to a single lymph node. N2b: The present of multiple lymph nodes. N2c: There are lymph nodes in the neck either on the opposite side as the main cancer or on both sides. N3: There is spread to one or more neck lymph nodes	N0 = 0; N1 = 1; N2a = 2; N2b = 3; N2c = 4; N3 = 5;	Numeric
Derived AJCC M, 7th edition (2010+) stage	M0; No distant metastasis M1; Distant metastasis	M0 = 0; M1 =1	Numeric
Radiation	Indication of whether patient has received radiation	0 = None, 1 = exposed to radiation	Numeric
Surgical resection	This describes if surgery was performed	0 = No surgery performed; 1 = Surgery not performed	Numeric
Overall survival	The time from the beginning of treatment until the last follow-up time or death	0 = Alive; 1 = Dead	Discrete

25

Table 3. The performance metrics of the cross-validated machine learning algorithms on the training data.

Algorithms	Accuracy (%)	AUC	Precision	F1 score	Recall
Logistic Regression	69.6	0.76	0.71	0.69	0.66
Naive Bayes	69.6	0.76	0.71	0.67	0.67
Support Vector Machine	69.5	0.76	0.71	0.68	0.66
Neural Network	73.1	0.83	0.78	0.70	0.63
Boosted Decision Tree	83.1	0.90	0.82	0.83	0.85
Decision Forest	81.5	0.89	0.81	0.82	0.83
Decision Jungle	79.6	0.88	0.80	0.79	0.79

Area Under Receiving Operating Characteristic (ROC) Curve (AUC); Recall = Sensitivity.

Table 4. The performance metrics of the comparison between the nomogram and machine learning model

Parameters	Nomogram (with surgical treatment)	Machine learning model	Nomogram (with radiotherapy)
True positive	6	6	6
False positive	18	6	21
True negative	29	41	26
False negative	0	0	0
Sensitivity	1	1	1
Specificity	0.62	0.87	0.55
F1 score	0.40	0.66	0.36
Accuracy (%)	**66.0**	**88.7**	**60.4**