Journal of Technology Research (JTR)
Volume 2, Issue 2, (2024), pp 053-058, ISSN 3005-639X

Machine Learning in Head and Neck Cancer: Clinical Prospects and Future Directions

Rasheed O. Alabi[1,2*], Alhadi Almangush[3,4], Mohammed Elmusrati[2]

[1]Research Program in Systems Oncology, Faculty of Medicine, University of Helsinki, Helsinki, Finland.
[2]Department of Industrial Digitalization, School of Technology and Innovations, University of Vaasa, Finland.
[3] Department of Pathology, University of Helsinki, Helsinki, Finland.
[4]University of Turku, Institute of Biomedicine, Pathology, Turku, Finland.

*Corresponding author email: rasheed.alabi@helsinki.fi

Received: 09-10-2024 | Accepted: 24-10-2024 | Available online: 15-12-2024 | DOI:10.26629/jtr.2024.07

ABSTRACT

The American Joint Committee on Cancer (AJCC) Tumor-Nodal-Metastasis (TNM) staging system has been widely used for planning of treatment strategies. However, for an individual patient, it is ineffective for predicting outcome due to its inability to consider other tumor- and patient-related risk factors. To this end, a tool that considers these factors together to accurately predict patients' outcomes would be pertinent. Objectives: This study aimed at examining the potential of a collaborative machine learning (cML)-based approach in estimating the overall survival of oral cancer patients. We compared the performance of cML with voting ensemble-based machine learning model. The prognostic significance of the clinicopathological parameters used to develop the model was examined using permutation feature importance. Furthermore, we examined some of the factors that can hinder the recommendation of machine learning (ML) models for further clinical evaluations. Methodology: The clinicopathological information of 9439 oral cancer patients were extracted from the Surveillance, Epidemiology, and End Results database, United States. Altogether, three machine learning (ML) models – voting ensemble, stacked ensemble, and extreme gradient boosting were combined to form a cluster of cML models. The performance of the cML was compared with the best performing individual ML algorithm in terms of accuracy. Results: The performance accuracy of voting ensemble, stacked ensemble, and extreme gradient boosting was 70.2%, 69.2%, and Place figures are at the bottom, while the tables are at the top, with the necessity of sequencing the numbering of the figures and tables as shown in TABLE 1. and Fig. 1. below 69.6%, respectively. When the cML and voting ensemble were randomly validated with 50 cases from the temporal validation cohort, they showed comparable performance. In terms of future importance, the most significant features were age of the patient at diagnosis, T stage, tumor grade, marital status, gender, primary site, surgery, N stage, radiation treatment, ethnicity, chemotherapy, and M stage. Discussion and Conclusions: The idea of the cML is to consider the unique properties of each of the ML models in making final predictions. Thus, representing a paradigm shift from individualism to cooperativism in the quest for personalized estimation of outcome for oral cancer patients. Of note, rather than competition among participating models, they cooperate to offer reliability to the final prediction made by the cML. Lack of independent geographic validation, model generalization and rigidity, and explainability are some of the factors identified that limit the recommendation of ML models for further clinical evaluation. The cML approach may aid in determining individualized treatment for oral cancer patients.

Keywords: Machine learning (ML); Head and Neck Squamous Cell Carcinoma (HNSCC); Overall survival; Explainability.

التعلم الآلي في سرطانات الرأس والرقبة: الآفاق السريرية والتوجهات المستقبلية

رشيد أوموبولاجي ألبي[2,1]، الهادي المنقوش[4,3]، محمد المصراتي[2]

[1]برنامج البحوث في علم الأورام المنظومي، كلية الطب، جامعة هلسنكي، هلسنكي، فنلندا.

[2]قسم الرقمنة الصناعية، كلية التكنولوجيا والابتكارات، جامعة فاسا، فنلندا.

[3]قسم علم الأمراض، جامعة هلسنكي، هلسنكي، فنلندا.

[4]جامعة توركو، معهد الطب الحيوي، قسم علم الأمراض، توركو، فنلندا.

ملخـــــــص البحـــــــث

لقد تم اســـتخدام نظام التدريج (Tumor–Nodal–Metastasis, TNM) التابع للجنة الأمريكية المشـــتركة للســـرطان (AJCC) على نطاق واســع في تخطيط اســتراتيجيات العلاج. إلا أن هذا النظام، بالنسـبة للمريض الفرد، غير فعّال في التنبؤ بالنتائج نظرًا لعجزه عن مراعاة عوامل الخطر الأخرى المرتبطة بالورم والمريض. ومن هذا المنطلق، يصـــبح من الضـــروري توفر أداة تراعي هذه العوامل مجتمعة من أجل التنبؤ بدقة بنتائج المرضـــى. هدفت هذه الدراســـة إلى فحص إمكانات النهج التعاوني القائم على التعلم الآلي (cML) في تقدير البقاء الكلي لمرضى سرطان الفم. كما تمت مقارنة أداء (cML) مع نموذج التعلم الآلي القائم على التصـــويت الجماعي (voting ensemble). وتم تحليل الأهمية التنبؤية للمعايير السريرية والمرضية المستخدمة في تطوير النموذج عبر طريقة Permutation Feature Importance. علاوة على ذلك، تمت دراســـة بعض العوامل التي قد تعيق التوصـــية باســـتخدام نماذج التعلم الآلي في التقييمات الســـريرية المستقبلية. تم استخراج البيانات السريرية والمرضية لـــ 9439 مريضًا بسرطان الفم من قاعدة بيانات Surveillance, Epidemiology, and End Results (SEER) بـالولايـات المتحـدة. وقد جرى دمج ثلاثـة نمـاذج للتعلم الآلي – التصـــويت الجمـاعي (Voting Ensemble)، والتكـديس الجمـاعي (Stacked Ensemble)، والتعزيز التـدريجي المتطرف (Extreme Gradient Boosting) – لتشـــكيل مجموعة من نماذج (cML). تمت مقارنة أداء (cML) مع أفضـل خوارزمية فردية من حيث الدقة، النتائج: بلغت دقة الأداء لنماذج التصـويت الجماعي والتكديس الجماعي والتعزيز التدريجي المتطرف 70.2%، 69.2%، و69.6% على التوالي. وعند التحقق العشـوائي من (cML) ونموذج التصـويت الجماعي باسـتخدام 50 حالة من مجموعة التحقق الزمني، أظهرا أداءً متقاربًا. أما من حيث أهمية الخصـائص، فقد تبين أن العوامل الأكثر تأثيرًا هي: عمر المريض عند التشـــخيص، مرحلة الورم (T stage)، درجة الورم، الحالة الاجتماعية، الجنس، الموقع الأولي للورم، التدخل الجراحي، مرحلة العقد اللمفاوية (N stage)، العلاج الإشـــعاعي، الأصـــل العرقي، العلاج الكيميائي، ومرحلة النقائل (M stage)، تقوم فكرة (cML) على الاسـتفادة من الخصـائص الفريدة لكل نموذج من نماذج التعلم الآلي في عملية التنبؤ النهائية، مما يمثل تحولًا من الفردية إلى التعاون في السعي نحو تقدير شخصي لنتائج مرضـى سرطان الفم. ومن الجدير بالذكر أن هذه النماذج، بدلًا من أن تتنافس، تتعاون لتقديم تنبؤ أكثر موثوقية. غير أن غياب التحقق الجغرافي المسـتقل، ومحدودية التعميم والمرونة، وضـعف القدرة على التفسـير، تُعد من أبرز العوامل التي تحد من التوصـية باعتماد نماذج التعلم الآلي للتقييمات السـريرية المسـتقبلية. ومع ذلك، قد يسـهم نهج (cML) في تحديد خطط علاجية فردية لمرضى سرطان الفم.

الكلمات الدالة: التعلم الآلي (**ML**)؛ سرطان الخلايا الحرشفية في الرأس والرقبة (**HNSCC**)؛ البقاء الكلي؛ قابلية التفسير..

1. INTRODUCTION

Oral squamous cell carcinoma (OSCC) represents the most frequent subsite of head and neck cancer [1–3]. The rates for incidence, recurrence, and mortality of OSCC have shown a marked increase in recent decades in the Western world due to the aggressive nature of this type of cancer in terms of its rapid local invasion and early lymph node metastasis [4–7]. Therefore, a concise effort is needed to predict OSCC tumor behavior, but the lack of specific prognostic indicators still constitutes a major challenge [4]. In addition, the decision-making regarding the best treatment approach is somewhat challenging for many cases of OSCC despite the general improved overall survival of OSCC patients.

Remarkably, OSCC is usually diagnosed late. This makes it relatively challenging to properly manage OSCC due to the experience of higher rates of recurrence and poor survival despite the recent improvements in OSCC diagnostic and management approaches [8,9]. Additionally, the treatment options for these patients may contribute to significant morbidity and psychosocial concerns. Late-stage OSCC is characterized by a significant burden on the patient's physical appearance (i.e. disfiguration), proper functioning (mastication and deglutition), major senses, airway, upper gastrointestinal tract, and low self-esteem in terms of social interactions and normal daily activities [8]. Thus, it is important to properly examine the patients to plan a targeted individualized treatment option by examining the 5-year overall survival (OS) prognosis.

2. MATERIALS AND METHODS

2.1 Collection of Data

In this study, we used the data from the National Cancer Institute (NCI) through the Surveillance, Epidemiology, and End Results (SEER) Program of the National Institutes of Health (NIH). This source was considered because it contains a large number of cases that can support large-scale analysis.

2.2 Ethical Permission

The ethical permission to use the SEER database of the NCI was granted with the identification number 17247-Nov2020 (alabir) for the specialized dataset. An extension to access the treatment-related parameters of the patients was granted with the same identification number.

2.3 Variable Selection

The included clinicopathological parameters available were age at diagnosis, ethnicity, gender, marital status, tumor grade, and stage classification according to the American Joint Committee on Cancer (AJCC) tumor-nodal-metastasis (TNM) 7th edition, and treatment parameters. Overall survival was the primary endpoint and target variable. The query of the SEER database produced a total of 9439 pathologically confirmed OSCC patients. Some of the selected variables were changed to categorical parameters and normalized for the ML training phase (sub-section 2.4)..

2.4 Machine Learning Training

In this project, several ensemble ML algorithms were selected. We chose ensemble and tree-based ML algorithms due to their promising results in cancer prognostication tasks [12,13]. Therefore, the selected ensemble algorithms were voting ensemble, stacked ensemble, extreme gradient boosting (XGBoost), light gradient boosting (light GBM), and logistic regression. We combined the voting ensemble, stacked, and Light GBM to form a collaborative predictive model (cML). The schematic for the collaborative paradigm is presented in Figure 1. The performance of the cML was compared with the highest-performing individual ensemble method. The extracted data (sub-section 2.3) were exported to Microsoft Azure Machine Learning for model training. Following the loading of data, we used a 5-fold cross-validation and adjusted other training parameters to guarantee better performance accuracy. In demonstrating the cML paradigm, a democratic voting approach was considered.

2.5. Performance Metrics of the Trained Model

The performance of the trained model was evaluated primarily using accuracy. Other performance metrics such as sensitivity, specificity, FI- score, and area under receiving operating characteristics curve were considered.

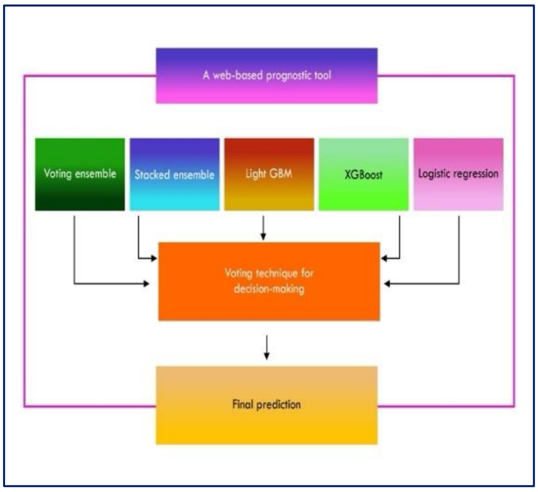

Fig 1. Collaborative paradigm.

3. RESULTS AND DISCUSSION

Following model training, the individual performance accuracy of the individual participating algorithms was 70.2%, 69.9%, 69.1%, 69.4%, and 69.5% for voting ensemble, stacked ensemble, XGBoost, Light GBM, and logistic regression, respectively. When the predictive outcomes of three of these algorithms were combined for collaborative decision-making (cML), the overall performance accuracy of the cML showed comparable performance with the voting ensemble. The feature importance of the input variables showed that the age of the patient at diagnosis, T stage, tumor grade, marital status, gender, primary site, surgery, N stage, radiation treatment, ethnicity, chemotherapy, and M stage, in decreasing order of importance, were significant for the model's ability to predict the overall survival of oral cancer patients Fig 2.

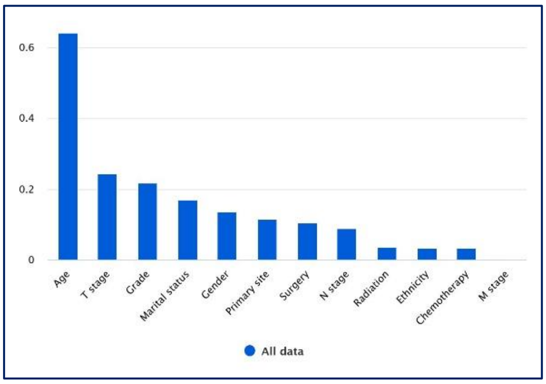

Fig 2. Significance of attributes.

The main limitations and concerns can be grouped as either the challenges inherent to the science of machine learning or relating to clinical implementations. The concern inherent to the science of machine learning includes the black-box concern, amount and quality of the data used in the training, unintended fitting of cofounders as input variables, and generalizability of the model (the predictive model can be used outside the data on which it was trained initially). The concerns relating to clinical implementation include interpretability and explainability, changing the fiducial relationship between the patient and clinicians, super-human analogy, and job-competitor

4. CONCLUSIONS

Individualized therapeutic decision-making based on survival prognosis remains challenging in the management of patients with OSCC. In this study, we have used an ML approach for overall survival (OS) prognosis. It is hoped that this model can assist clinicians to make an informed decision regarding future treatment options. The individualized targeted treatment can prevent overtreatment of OSCC cancer patients, thereby, improving their quality of health (QoH), and quality of life (QoL). In this study, we leveraged a collaborative ML (cML) paradigm so that rather than considering the model as a single entity (model individualism), we combined the unique properties of each ML algorithm to form model cooperativism. This study showed that cML showed comparable performance with the highest-performing ensemble method. Of note, the idea of cML may not necessarily be hinged on performance enhancements. However, considering the sensitive nature of medical applications such as cancer management where the reliability of the prediction is pertinent, a collaborative (cML) approach becomes warranted. Our model highlighted the age of the patient at diagnosis, and the T stage as the top prognostic parameters for OS prediction in OSCC. This finding is supported previous findings in the literature [14–17].

Remarkably, the approach of cML also seeks to address some of the concerns limiting recommending ML models for further clinical evaluations. In recent years, several efforts have been made to address concerns relating to the model's interpretability and explainability using both SHapley Additive exPlanations (SHAP) and Local Interpretable Model Agnostic Explanations (LIME) techniques [18,19]. Similarly, continuous efforts are being made to further validate ML models using independent external validations to facilitate model generalizability [19,20]. While model generalizability reveals the performance of the model with independent data outside the training cohorts, efforts aimed at continuous model improvement are pertinent. The idea of having web-based prognostic may not address continuous model development, rather it seeks to further validate the developed models. Therefore, to benefit from the variability in the data used for either temporal validation or independent geographic validation, it is important to explore other paradigms without comprising data security and privacy-related issues. An example of such a paradigm is the use of federated machine learning [21].

With the advancements in technology and availability of medical data in various formats such as computed tomography (CT), positron emission tomography (PET), and magnetic resonance imaging (MRI), further diagnosis and prognosis using a modified artificial neural network, that is, a convolutional neural network (CNN) have been explored in recent years [22]. The deep learning (DL) approach is aimed at ensuring personalized medicine from radiological images of cancer patients. In recent years, insightful parameters have been explored from these images through radiomics [23]. Radiomics extracted features are now combined with clinicopathological features or genomics parameters to further enhance a robust model that can facilitate personalized oncology [24].

In conclusion, our study further emphasized the potential of ML for outcome prognostication and personalized medicine to improve OSCC management. Despite the promising results showing the potential of ML for OSCC management, these models should be developed to further enhance explainability, interpretability, and externally validated for generalizability in order to be safely integrated into daily clinical practices. Also, regulatory frameworks for the adoption of these models in clinical practices are necessary. Our study has some limitations. First, the data used for ML model development were retrospective in nature. Second, the models are not externally validated to evaluate the true performance of the model. In future studies, it is essential to further validate the cML and individual ML model using a relatively large amount of data to further evaluate the potential of cML. This is important to further fulfill our ultimate goal of providing a reliable prediction from an ML model that can aid in personalized treatment plans for patients with OSCC.

5. REFERENCES

[1] Marur S, Forastiere AA. Head and neck squamous cell carcinoma: update on epidemiology, diagnosis, and treatment. Mayo Clin Proc. 2016;91(3):386-96. doi:10.1016/j.mayocp.2015.12.017.

[2] Pires FR, Ramos AB, Oliveira JBC, Tavares AS, Luz PSR, Santos TCRB. 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(5):460-7. doi:10.1590/1679-775720130317

[3] Li R, Koch WM, Fakhry C, Gourin CG. Distinct epidemiologic characteristics of oral tongue cancer patients. Otolaryngol Head Neck Surg. 2013;148(5):792-6. doi:10.1177/0194599813477992

[4] Prognostic factors in tongue cancer – relative importance of demographic, clinical and histopathological factors. Br J Cancer. 2000;83(5):614-9. doi:10.1054/bjoc.2000.1323

[5] Bello IO, Soini Y, Salo T. Prognostic evaluation of oral tongue cancer: means, markers and perspectives (I). Oral Oncol. 2010;46(9):630-5. doi:10.1016/j.oraloncology.2010.06.006

[6] 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 Arch. 2015;467(1):39-46. doi:10.1007/s00428-015-1758-z
[7] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209-49. doi:10.3322/caac.21660
[8] Mroueh R, Haapaniemi A, Saarto T, Grönholm L, Grénman R, Salo T, et al. Non-curative treatment of patients with oral tongue squamous-cell carcinoma. Eur Arch Otorhinolaryngol. 2019;276(7):2039-45. doi:10.1007/s00405-019-05456-y
[9] Alabi RO, Mäkitie AA, Pirinen M, Elmusrati M, Leivo I, Almangush A. Comparison of nomogram with machine learning techniques for prediction of overall survival in patients with tongue cancer. Int J Med Inform. 2021;145:104313. doi:10.1016/j.ijmedinf.2020.104313
[10] Alabi RO, Elmusrati M, Sawazaki-Calone I, Kowalski LP, Haglund C, Coletta RD, et al. Machine learning application for prediction of locoregional recurrences in early oral tongue cancer: a web-based prognostic tool. Virchows Arch. 2019;475(4):489-97. doi:10.1007/s00428-019-02642-5
[11] Alabi RO, Elmusrati M, Sawazaki-Calone I, Kowalski LP, Haglund C, Coletta RD, et al. Comparison of supervised machine learning classification techniques in prediction of locoregional recurrences in early oral tongue cancer. Int J Med Inform. 2019;129:104068. doi:10.1016/j.ijmedinf.2019.104068
[12] Mäkitie AA, Alabi RO, Ng SP, Takes RP, Robbins KT, Ronen O, et al. Artificial intelligence in head and neck cancer: a systematic review of systematic reviews. Adv Ther. 2023;40(8):3360-80. doi:10.1007/s12325-023-02527-9.
[13] Alabi RO, Youssef O, Pirinen M, Elmusrati M, Mäkitie AA, Leivo I, et al. Machine learning in oral squamous cell carcinoma: current status, clinical concerns and prospects for future—a systematic review. Artif Intell Med. 2021;115:102060. doi:10.1016/j.artmed.2021.102060
[14] Pai AY, Rizvi A, Bryant DG, Mitsimponas K. UICC TNM8 implementation and its effects on head and neck cancer staging in a tertiary referral centre. Adv Oral Maxillofac Surg. 2021;1:100022. doi:10.1016/j.adoms.2021.100022
[15] Adoga AA, Kokong DD, Ma'an ND, Mugu JG, Mgbachi CJ, Dauda AM. The predictive factors of primary head and neck cancer stage at presentation and survival in a developing nation's tertiary hospital. SAGE Open Med. 2018;6:205031211879241. doi:10.1177/2050312118792416
[16] Cadoni G, Giraldi L, Petrelli L, Pandolfini M, Giuliani M, Paludetti G, et al. Prognostic factors in head and neck cancer: a 10-year retrospective analysis in a single-institution in Italy. Acta Otorhinolaryngol Ital. 2017;37(6):458-66. doi:10.14639/0392-100X-1246
[17] Talani C, Mäkitie A, Beran M, Holmberg E, Laurell G, Farnebo L. Early mortality after diagnosis of cancer of the head and neck: a population-based nationwide study. PLoS One. 2019;14(10):e0223154. doi:10.1371/journal.pone.0223154
[18] Alabi RO, Elmusrati M, Leivo I, Almangush A, Mäkitie AA. Machine learning explainability in nasopharyngeal cancer survival using LIME and SHAP. Sci Rep. 2023;13:8984. doi:10.1038/s41598-023-35795-0
[19] Alabi RO, Almangush A, Elmusrati M, Leivo I, Mäkitie AA. An interpretable machine learning prognostic system for risk stratification in oropharyngeal cancer. Int J Med Inform. 2022;168:104896. doi:10.1016/j.ijmedinf.2022.104896
[20] Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where? Clin Kidney J. 2021;14(1):49-58. doi:10.1093/ckj/sfaa188
[21] Sheller MJ, Edwards B, Reina GA, Martin J, Pati S, Kotrotsou A, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Sci Rep. 2020;10:12598. doi:10.1038/s41598-020-69250-1
[22] Alabi RO, Almangush A, Elmusrati M, Mäkitie AA. Deep machine learning for oral cancer: from precise diagnosis to precision medicine. Front Oral Health. 2022;2:794248. doi:10.3389/froh.2021.794248
[23] Alabi RO, Elmusrati M, Leivo I, Almangush A, Mäkitie AA. Artificial intelligence-driven radiomics in head and neck cancer: current status and future prospects. Int J Med Inform. 2024;188:105464. doi:10.1016/j.ijmedinf.2024.105464
[24] Bruixola G, Remacha E, Jiménez-Pastor A, Dualde D, Viala A, Montón JV, et al. Radiomics and radiogenomics in head and neck squamous cell carcinoma: potential contribution to patient management and challenges. Cancer Treat Rev. 2021;99:102263. doi:10.1016/j.ctrv.2021.102263