Interpretable machine learning model for prediction of overall survival in laryngeal cancer

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

Title: Interpretable machine learning model for prediction of overall survival in laryngeal cancer

Year: 2024

Version: Accepted manuscript

Interpretable machine learning model for prediction of overall survival in laryngeal cancer

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

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

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

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

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

[e] Department of Otorhinolaryngology – Head and Neck Surgery, University of Helsinki and Helsinki University Hospital, Helsinki, Finland.

[f] Division of Ear, Nose and Throat Diseases, Department of Clinical Sciences, Intervention and Technology, Karolinska Institute and Karolinska University Hospital, Stockholm, Sweden.

Corresponding Author: Antti A. Mäkitie
Research Program in Systems Oncology, Faculty of Medicine, University of Helsinki, Helsinki, Finland. **E-mail address:** antti.makitie@helsinki.fi

Disclosure: The authors declare no conflicts of interest.

Abstract

Background: The mortality rates of laryngeal squamous cell carcinoma cancer (LSCC) have not significantly decreased in the last decades. **Objectives**: We primarily aimed to compare the predictive performance of DeepTables with the state-of-the-art machine learning (ML) algorithms (Voting ensemble, Stack ensemble, and XGBoost) to stratify patients with LSCC into chance of overall survival (OS). In addition, we complemented the developed model by providing interpretability using both global and local model-agnostic techniques. **Methods:** A total of 2792 patients in the Surveillance, Epidemiology, and End Results (SEER) database diagnosed with LSCC were reviewed. The global model-agnostic interpretability was examined using SHapley Additive exPlanations (SHAP) technique. Likewise, individual interpretation of the prediction was made using Local Interpretable Model Agnostic Explanations (LIME). **Results:** The state-of-the-art ML ensemble algorithms outperformed DeepTables. Specifically, the examined ensemble algorithms showed comparable weighted area under receiving curve of 76.9, 76.8, and 76.1 with an accuracy of 71.2%, 70.2%, and 71.8%, respectively. The global methods of interpretability (SHAP) demonstrated that the age of the patient at diagnosis, N stage, T stage, tumor grade, and marital status are among the prominent parameters. **Conclusions:** A ML model for OS prediction may serve as an ancillary tool for treatment planning of LSCC patients.

KEYWORDS: Machine learning; Deep learning; DeepTables; Laryngeal cancer; Laryngeal Squamous Cell Carcinoma; XGBoost; Voting ensemble; Stacked Ensemble; SEER; Overall survival

Introduction

Laryngeal squamous cell carcinoma (LSCC) was expected to affect 12,470 (0.7% of all new cancer diagnoses) new cases and lead to 3,820 (0.6% of all cancer mortality) deaths in 2022 [1]. The marked decrease (about 2.4% yearly decrement) in the annual incidence of LSCC is mainly due to the continuous decrease in the use of tobacco products and excessive consumption of alcohol [1]. Despite the incidence decrement, little or no change has been observed in the 5-year overall survival of LSCC patients for the past years [2]. The chance of survival depends on several factors such stage at diagnosis. Remarkably, a significant number of LSCC are diagnosed at a late stage with the chance of survival having decreased to as low as 40% [1]. Therefore, it is important to properly develop and plan for overall management of LSCC and thus, to improve the chance of survival.

The treatment paradigm of LSCC has evolved due to poor quality of life and deterioration in important functions that involve swallowing, voice, and breathing because of the sequelae of treatment - surgery, radiotherapy, chemotherapy, and multimodality treatment. Therefore, it becomes important to avoid overtreatment and undertreatment of LSCC patients to improve their quality of life without sacrificing survival rates. Disease extent and anatomic parameters warrants careful consideration of the treatment approach [1]. An important and insightful approach in achieving this is to stratify the patients into risk groups to ensure a targeted treatment approach [3]. In recent years, machine learning (ML) which is a subfield of artificial intelligence (AI) has shown a promising ability as an ancillary tool to aid clinicians in the overall survival (OS) risk stratification of various cancers [4–7]. In recent years, another subfield of AI, deep learning (DL), has shown promising results in medical image analysis aimed at improving management of cancer [8–10]. However, for tabular data, DL has arguably found little or no usage for developing diagnostic/prognostic models for personalized oncology.

Based on the recent success of artificial intelligence (AI) in health care [10–12], we aimed to examine the potential of ML in the prediction of overall of LSCC patients for careful treatment planning. We primarily aimed to explore the potential of DeepTables, which is a DL approach on tabular data for overall survival prediction in LSCC. Secondarily, we aimed to compare the performance of DeepTables to the state-of-the-art ensemble approaches – voting ensemble, stacked machine learning, and extreme gradient boosting ML algorithms. In addition, we explored the contributions of each variable to the overall performance of the predictive model using SHapley Additive exPlanations (SHAP) techniques. Similarly, the local interpretation, i.e., the contribution of each variable for the prediction of an instance was examined using the Local Interpretable Model Agnostic Explanations (LIME).

2. Material and Methods

2.1. Study endpoint

The endpoint in this study was the OS of LSCC patients. This endpoint was stratified into two groups – alive or dead. Alive as an endpoint of overall survival means that the trained model will predict the patient to be alive while dead as an endpoint of overall survival indicates that the trained ML model predicts the patient to die during the follow-up and considering all causes of death.

2.2. Dataset creation/curation and patients' attributes selection

All patients were extracted from the Surveillance, Epidemiology, and End Results (SEER) program by selecting the Research Plus Data for 8 Registries, Nov 2022 Sub (1975 – 2020). The permission to use the SEER database was approved by identification number 17247-Nov2020. The inclusion criteria were: (1) histologically confirmed SCC of the larynx; (2) known OS status at the end of follow-up; (3) known follow-up time in months. Conversely, the exclusion criteria were: (1) Other cancer subsites (2) patients with incomplete data records or "lost to follow-up"; and (3) patients with unknown overall survival status. Following this inclusion and exclusion criteria, a total of 2792 LSCC patients were curated. Clinical and pathological parameters were recorded as follows: year of diagnosis, age at diagnosis (No cut off was used), ethnicity, gender, marital status, grade, and stage classification according to the American Joint Committee on Cancer (AJCC) tumor-nodal-metastasis (TNM) 7th edition (Table 1). The included treatment parameters were surgery, radiotherapy, and chemotherapy. OS was the primary endpoint and target variable (Table 1). These extracted parameters were in conformity with a similar study where these parameters were considered significant [13].

2.3. Dataset pre-processing

Following the dataset curation process (**Section 2.2**), a series of preprocessing operations were carried out on the extracted raw data to optimize the information extraction process by identifying missing values. All cases of missing values were deleted from the extracted data. String parameters were converted into categorical variables for model development (**Section 2.4**).

2.4. Machine learning model training

We used Python 3.11.4 and Microsoft Machine Learning Azure for model training. We used Python for 3.11.4 for the training of the DeepTables while the ensemble models (voting, stacked, and extreme gradient boosting) were developed in Microsoft Machine Learning Azure. Following the data preprocessing (**Section 2.3**), the processed data were loaded into Jupyter Notebook using Python version 3.11.4. The necessary dependencies for the DeepTables were imported. The extracted data were randomly divided into 50:50 training and testing set for DeepTables model development (Figure 1). Various hyperparameters were tuned to ensure that the model's performance is reasonable. The DeepTables model was evaluated based on accuracy.

Similarly, the ensemble machine learning models were developed using a 5-fold cross-validation. Following data importation, attributes (inputs and output) definition, algorithm selection, hyperparameters tunning, the model was trained. The performance of the trained model was primarily evaluated based on its accuracy. Following the model training phase, the hyperparameters were further tuned to maximize the performance of the model (Figure 1). Other performance metrics of the ensemble method are given in Table 2.

2.5. Model interpretation

We explored the model's interpretation from two distinct perspectives. First, we examined the contribution of each variable for the prediction of an instance (local interpretation) using the Local Interpretable Model Agnostic Explanations (LIME). Second, we explored the contributions of each variable to the overall performance of the predictive model (global interpretability) using SHapley Additive exPlanations (SHAP) techniques. Both LIME and SHAP were examined on the algorithm that showed the best predictive accuracy.

3. Results

3.1. Patient characteristics

The patient-, tumor- and treatment-related characteristics of the extracted cohort are summarized in Table 1. Many of the patients in our series were middle-aged and elderly men over 40 years of age (Table 1). As shown in Table 1, 2261 (81.0%) were above 40 years of age. In general, the average age at diagnosis for LSCC patients included in this study was 65.3 years (range, 21– 90; median age at diagnosis of 65.0; SD $\pm$ 11.1). The male-to-female ratio was 4.3:1 (81.0% male, 19.0% female). In terms of ethnicity, 2325 (83.3%) were Caucasian, 285 (10.2%) were black, and 182 (6.5%) were from other origins including American Indian/AK Native, Asian/Pacific Islander. The marital status of LSCC revealed that 1432 (51.3%) were married while 1360 (54.4%) were considered unmarried (single, divorced, widowed, and separated) at the time of diagnosis (Table 1).

Considering the TNM classification, 887 (31.8%) had T1, 775 (27.8%) had T2, 674 (24.1%) had T3, and 456 (16.3%) had T4. Regarding the N-stage, N0 covered a total of 1940 (69.5%) patients, while 288 (10.3%) had N1 and 533 (19.1%) had N2 and 31 (1.1%) had N3. A significant number of patients (97.4%) had M0. In terms of tumor grading, 464 (16.6%) tumors were well differentiated, 1648 (59.0%) moderately differentiated, 656 (23.5%) were poorly differentiated, and 24 (0.86%) were undifferentiated tumors. The histopathologic characteristics are briefly summarized in Table 1. Three different treatment options were provided for the LSCC. A total of 2160 (77.4%) patients received radiotherapy treatment.

3.3. Model performance

The ensemble algorithms – voting ensemble, stack ensemble, and XGBoost showed a weighted area under receiving curve of 76.9, 76.8, and 76.1 with an accuracy of 71.2%, 70.2%, and 71.8%, respectively. Conversely, the DeepTables showed an accuracy of 59.7%. This shows that the ensemble algorithms clearly outperformed the DeepTables algorithm. Although, the ensemble algorithms showed comparatively similar predictive performances.

3.4. Model variable performance

As shown in Figure 2, the local interpretation of the prediction made by the model showed that the degree of accuracy (prediction probabilities) of the prediction was 52.0% for the LSCC to have a poor overall survival and 48.0% for good overall survival. In addition, for this particular instance of prediction, the age of the patients at diagnosis and marital status contributed to the prediction (dead as the endpoint of overall survival) made by the model. Similarly, in terms of the global interpretation of the model, that is, the contribution of each variable to the predictive ability of the model, the age of the patient at diagnosis, N-stage, T-stage, tumor grade, and marital status are among the most important parameters (Figure 3).

4. Discussion

We examined the prediction of overall survival (OS) of laryngeal cancer patients using the subfields of artificial intelligence (AI) – classic and deep machine learning paradigms. We provided interpretations to the predictions made by this model using SHapley Additive exPlanations (SHAP) and Local Interpretable Model Agnostic Explanations (LIME) techniques. We found that the ensemble methods (voting ensemble, stack ensemble, and extreme gradient boosting (XGBoost) outperformed the deep learning (DL) methodology (DeepTables) for tabular data. In addition, we found that the examined ensemble machine learning (ML) methods showed comparable performance. Furthermore, LIME provided the probability of the correctness of the prediction while SHAP explained how each parameter contributed to the overall predictive ability of the model.

DL has shown promising results in medical image analysis [8]. However, for tabular data, these methods have not been widely used. As shown in this study, tree-based ensemble models remain state-of-the-art for tabular data [14]. Our finding is corroborated by the findings of Grinsztajn et al. [15]. Several factors may have contributed to the better performance of the ensemble method over the DL method. For example, the size of the data [15]. Grinsztajn et al. further emphasized that tabular data are usually characterized by uninformative features, irregular patterns in the target function, and non-rotationally-invariant data in such a way that the linear combinations of features misrepresent the information [15]. In our study, we used a relatively small amount of data. It remains unclear how the deep learning approach would perform on a clearly larger dataset (typically more than 10,000 samples).

The development of a predictive model for the chance of OS prediction in laryngeal cancer would serve as an ancillary tool for the clinician to aid in decision-making. Considering the high mortality and morbidity associated with laryngeal cancer [2], having prior information about the chance of survival of LSCC can facilitate the development of a more personalized treatment plan for effective management of this neoplasm [16]. Having a prior information about the probability of overall survival can inform targeted treatment intervention. This can further guide the treatment decisions by the clinician. With this type of ancillary tool, clinician can further make insightful decisions that would spare patients from possibly ineffective treatment approaches. The predictive performance showed by the ML model presented in this study is capable of assisting clinicians in selecting a treatment approach that contributes to reducing morbidity, improving the chance of overall survival, and enhancing function and quality of life.

Despite the touted benefits of ML in oncology, several factors have impeded the recommendation of these models for further external validation and clinical evaluations [17]. Explainability and interpretability are among these factors [4]. In this study, the LIME paradigm provided the degree of correctness of the prediction. Most importantly, it highlights how each of the input variables contributed to the prediction made by the model and thus provides an interpretation of the prediction made by the model. The significance of this is that the clinicians using the model can clearly see how each of the input prognostic parameters contribute to the prediction made by the model.

Beyond the interpretation made by the model for a single patient, the SHAP approach explains how each variable contributed to the overall predictive performance of the model. In our study, the age of the patient at diagnosis, N stage, T stage, tumor grade, and marital status are among the prominent prognostic parameters for the predictive ability of the model. The study by Li et al. supported our finding that the age of the patients at diagnosis is an important prognostic factor since younger patients are likely to present less aggressive tumor biology than older ones [18]. Interestingly, this finding is further justified since a sizeable number of the patients in our series were above 40 years of age.

Presently, TNM staging forms the cornerstone of the treatment of laryngeal cancer [13]. This further explains why both T and N stages are among the prominent factors for the model's predictive ability. Tumor grading is an important prognostic factor because poorly differentiated tumors usually have a higher rate of metastatic disease. In addition, the accuracy of the degree of differentiation is usually affected by the subjectivity of interpretation by experienced pathologists [19]. Besides the T and N stages and tumor grade, marital status was found to be an important prognostic factor contributing to the predictive performance of the model. This is also supported by several previous studies where marital status has long been recognized as an important prognostic factor for many cancers [20]. Our finding was in tandem with a recent study by Ding et al. which emphasized that marital status is an independent factor affecting the prognosis of laryngeal cancer [20].

Our study has some limitations. We have used SEER that has certain inherent limitations. For example, the SEER database lacks important information of the patients such as parameters impacting OS such as performance status, comorbidities, smoking status and so on. In addition, there are limitations regarding the recording of the radiotherapy and chemotherapy information in the SEER registry. For example, there were no clear information regarding the exact radiation techniques. Another limitation was that the SEER database has no possibility to download the corresponding the image data of these patients. As a result, it was not feasible to perform radiomics analysis. Additionally, we did not compare the performance of our model with statistical analyses. Therefore, in the future, a ML model that combines important parameters such as comorbidities and smoking status with other clinicopathologic information is warranted. A radiomics-based model that incorporates radiomics generated features with clinicopathologic parameters would be beneficial for improved prediction of overall survival of laryngeal cancer patients. In the future, we foresee that our model would be enhanced with radiomics and radiogenomics-generated data. This enhanced model can be integrated as a web-based tool for improved performance, accessibility, and continuous model improvement. This can assist clinicians in selecting a targeted treatment approach that contributes to reducing morbidity and enhancing function and quality of life.

In conclusion, laryngeal cancer remains an economic and social burden to the healthcare system, caregivers, and patients. Considering the possibility of bias and subjectivity that may arise from radiologists, pathologists, surgeons, and oncologists, ML predictive models may serve as an auxiliary tool to aid in treatment planning and decision-making. In this study, we demonstrated that deep learning for tabular data did not outperform state-of-the-art ensemble methods despite its promising performance on text (language), image, or even audio datasets. Additionally, we developed ensemble decision-tree algorithms that were highly effective in predicting the chance of overall survival of laryngeal cancer. We provided the degree of correctness of the prediction made by the model and evaluated how each variable contributed to the overall predictive performance of the model. The present study used retrospective data and therefore prospective cohort is necessary. In addition, the developed models require further external independent validation.

Acknowledgment

The Sigrid Jusélius Foundation. State funding for the Helsinki University Hospital Research Fund. Finska Läkaresällskapet.

Disclosure statement

The authors report that there are no competing interests to declare.

Figure Legend

Figure 1. The schematic of machine learning training process.

Figure 2. The LIME interpretability for a prediction instance.

Figure 3. The SHAP interpretability on the contribution of each variable.

Tables

Table 1. Baseline demographic and tumor characteristics of laryngeal cancer patients extracted from SEER database.

Table 2. Performance metrics of the ensemble methods.

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

Variables (Definition)	**Total (N = 2792) (%)**	**Categorization for ML analysis**	**Data type after categorization**
Race Ethnicity of the patient			
White	2325 (83.3%)	0 = White	Numeric
Black	285 (10.2%)	1 = Black	
Others	182 (6.5%)	2 = Others (American Indian /AK Native, Asian pacific	
Age at diagnosis Age of the patient at diagnosis			
<40 years old (Young)	31 (1.1%)	No categorization for model training	Discrete
>=40 years old (Old)	2761 (98.9%)		
Gender Biological sex			
Female	531 (19.0%)	0 = Female	Numeric
Male	2261 (81.0%)	1 = Male.	
Marital Status Marital status of the patient at the time of diagnosis of NPC			
Single (never married)	1360 (48.7%)	0 = Single (never married)	Numeric
Married	1432 (51.3%)	1 = Married	
AJCC 7th edition, T-stage (2010 – 2015)			
AJCC T1	887 (31.8%)	T1 = 1	Numeric
AJCC T2	775 (27.8%)	T2 = 2	
AJCC T3	674 (24.1%)	T3 = 3	
AJCC T4	456 (16.3%)	T4 = 4	
AJCC 7th edition, N-stage (2010 – 2015)			
AJCC N0; No regional lymph node metastasis	1940 (69.5%)	N0 = 0	Numeric
AJCC N1; Single node regional lymph node metastasis	288 (10.3%)	N1 = 1	
AJCC N2; Cancer has spread to single lymph nodes	533 (19.1%)	N2 = 2	
AJCC N3; Cancer has spread to one or more lymph node	31 (1.1%)	N3 = 3	
AJCC 7th edition, M-stage (2010 – 2015)			
AJCC M0; No distant metastasis	2720 (97.4%)	M0 = 0	Numeric
AJCC M1; Presence of distant metastasis	72 (2.6%)	M1 = 1	
Grade The differentiation of cancer cell			
Grade I: Well differentiated	464 (16.6%)	Grade I = 1	Numeric
Grade II: Moderately differentiated	1648 (59.0%)	Grade II = 2	
Grade III: Poorly differentiated	656 (23.5%)	Grade III = 3	
Grade IV: Undifferentiated	24 (0.86%)	Grade IV = 4	
Surgical resection Indication of the performance of surgery			
No surgery performed	1581 (56.6%)	0 = No surgery performed	Numeric
Surgery performed	1211 (43.4%)	1 = Surgery performed	
Radiotherapy This describes whether the patient receives radiation or not			
No exposure to radiotherapy	632 (22.6%)	0 = No exposure to radiation	Numeric
Exposure to radiation therapy	2160 (77.4%)	1 = Exposure to radiation	
Chemotherapy			
No chemotherapy administered	1759 (63.0%)	0 = No chemotherapy administered	Numeric
Chemotherapy was administered	1033 (37.0%)	1 = Chemotherapy was administered	
Overall Status			
Alive	1113 (39.9%)	0 = Alive	Numeric
Dead	1679 (60.1%)	1 = Dead	

Table 2. Overview of the performance model of ensemble methods (voting ensemble, stack ensemble, and extreme gradient boosting [XGBoost]).

	Performance metrics	Voting ensemble	Stack ensemble	XGBoost
Confusion matrix parameters	True positive	664	498	664
	False positive	449	615	449
	False negative	339	216	339
	True negative	1340	1463	1340
Predictive value	PPV (Precision)	0.60	0.45	0.60
	NPV	0.80	0.87	0.80
Rate	False positive rate	0.25	0.30	0.25
	False negative rate	0.34	0.30	0.34
Other metrics	Sensitivity (recall)	0.66	0.69	0.66
	Specificity	0.75	0.70	0.75
	F1 score	0.63	0.55	0.63
Accuracy	Accuracy	71.8%	70.2%	71.8%
	Balanced accuracy	70.0%	66.0%	70.0%
Correlation	Mathew’s correlation coefficient	0.40	0.36	0.40
AUC	Weighted AUC	0.77	0.77	0.76

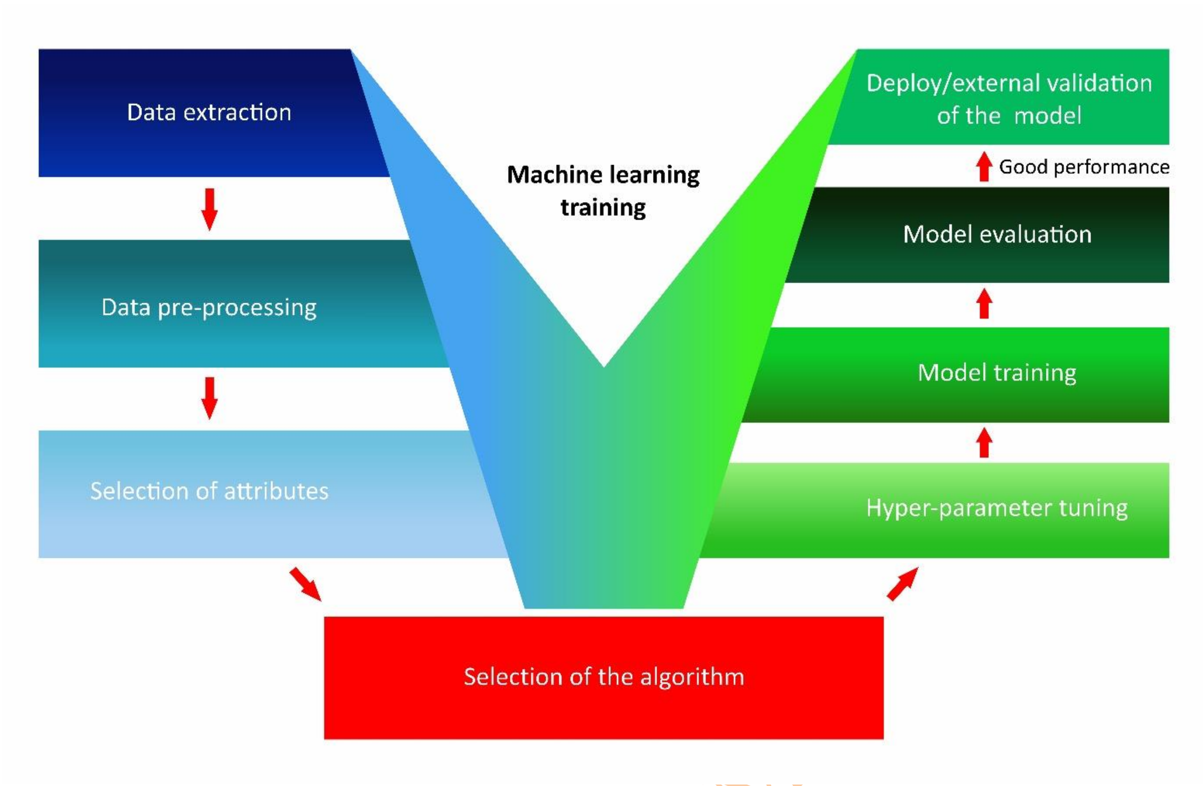

Figure 1. The schematic of machine learning training process.

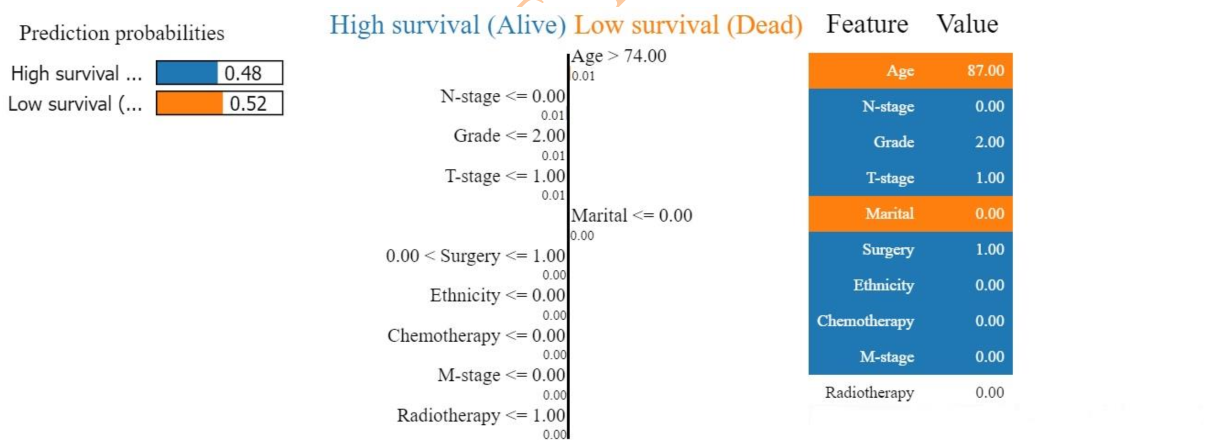

Figure 2. The LIME interpretability for a prediction instance.

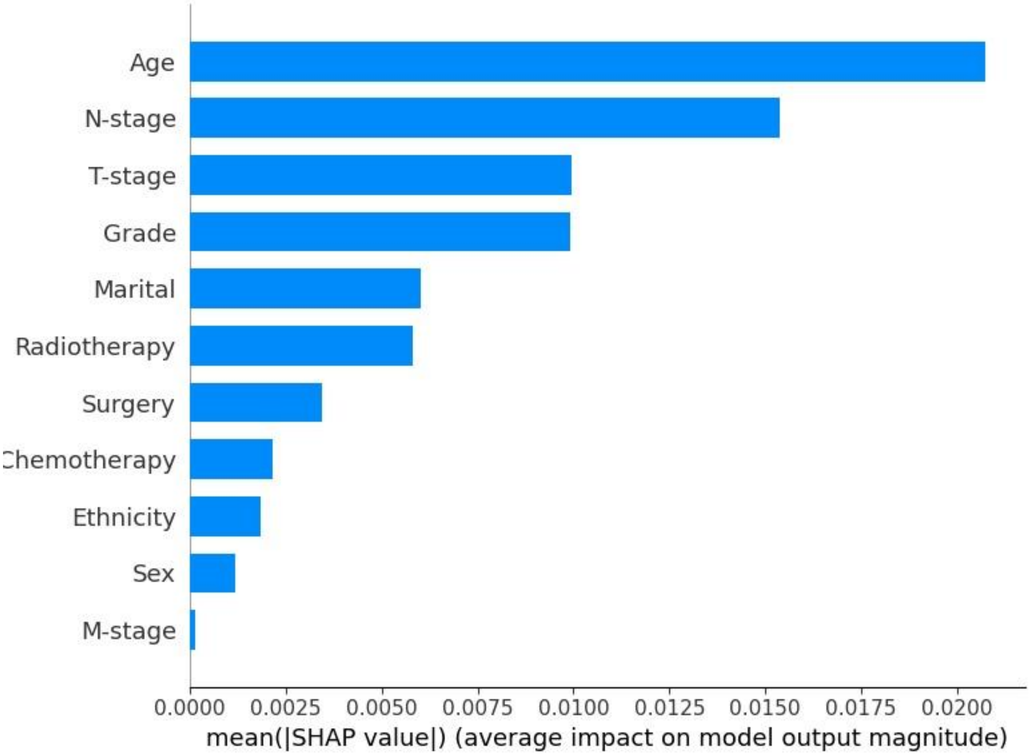

Figure 3. The SHAP interpretability on the contribution of each variable.

References

[1] Obid R, Redlich M, Tomeh C. The Treatment of Laryngeal Cancer. Oral and Maxillofacial Surgery Clinics of North America. 2019;31:1–11.

[2] Siegel RL, Miller KD, Fuchs HE, et al. Cancer statistics, 2022. CA A Cancer J Clinicians. 2022;72:7–33.

[3] Alabi RO, Elmusrati M, Sawazaki-Calone I, et al. Machine learning application for prediction of locoregional recurrences in early oral tongue cancer: a Web-based prognostic tool. Virchows Archiv. 2019;475:489–497.

[4] Alabi RO, Sjöblom A, Carpén T, et al. Application of artificial intelligence for overall survival risk stratification in oropharyngeal carcinoma: A validation of ProgTOOL. International Journal of Medical Informatics. 2023;175:105064.

[5] Alabi R, Almangush A, Elmusrati M, et al. An interpretable machine learning prognostic system for risk stratification in oropharyngeal cancer. International Journal of Medical Informatics. 2022;168:104896.

[6] Alabi RO, Elmusrati M, Leivo I, et al. Advanced-stage tongue squamous cell carcinoma: a machine learning model for risk stratification and treatment planning. Acta Oto-Laryngologica. 2023;1–9.

[7] Alabi RO, Elmusrati M, Leivo I, et al. Machine learning explainability in nasopharyngeal cancer survival using LIME and SHAP. Sci Rep. 2023;13:8984.

[8] Zhang Y-P, Zhang X-Y, Cheng Y-T, et al. Artificial intelligence-driven radiomics study in cancer: the role of feature engineering and modeling. Military Med Res. 2023;10:22.

[9] Alabi RO, Almangush A, Elmusrati M, et al. Deep Machine Learning for Oral Cancer: From Precise Diagnosis to Precision Medicine. Front Oral Health. 2022;2:794248.

[10] Alabi RO, Bello IO, Youssef O, et al. Utilizing Deep Machine Learning for Prognostication of Oral Squamous Cell Carcinoma—A Systematic Review. Frontiers in Oral Health [Internet]. 2021 [cited 2021 Jul 26];2. Available from: https://www.frontiersin.org/articles/10.3389/froh.2021.686863/full.

[11] Alabi RO, Elmusrati M, Sawazaki-Calone I, 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.

[12] Alabi RO, Mäkitie AA, Pirinen M, et al. Comparison of nomogram with machine learning techniques for prediction of overall survival in patients with tongue cancer. International Journal of Medical Informatics. 2021;145:104313.

[13] Petruzzi G, Coden E, Iocca O, et al. Machine learning in laryngeal cancer: A pilot study to predict oncological outcomes and the role of adverse features. Head & Neck. 2023;45:2068–2078.

[14] Kossen J, Band N, Lyle C, et al. Self-Attention Between Datapoints: Going Beyond Individual Input-Output Pairs in Deep Learning [Internet]. arXiv; 2022 [cited 2023 Aug 16]. Available from: http://arxiv.org/abs/2106.02584.

[15] Grinsztajn L, Oyallon E, Varoquaux G. Why do tree-based models still outperform deep learning on tabular data? [Internet]. arXiv; 2022 [cited 2023 Aug 16]. Available from: http://arxiv.org/abs/2207.08815.

[16] Kumar S, Patnaik S, Dixit A. Predictive models for stage and risk classification in head and neck squamous cell carcinoma (HNSCC). PeerJ. 2020;8:e9656.

[17] Alabi RO, Youssef O, Pirinen M, et al. Machine learning in oral squamous cell carcinoma: Current status, clinical concerns and prospects for future—A systematic review. Artificial Intelligence in Medicine. 2021;115:102060.

[18] Li R, Yu S, Zhu W, et al. Studying the impact of young age on prognosis and treatment in laryngeal squamous cell carcinomas using the SEER database. PeerJ. 2019;7:e7368.

[19] Bradford CR, Ferlito A, Devaney KO, et al. Prognostic factors in laryngeal squamous cell carcinoma. Laryngoscope Investigative Otolaryngology. 2020;5:74–81.

[20] Ding Z, Yu D, Li H, et al. Effects of marital status on overall and cancer-specific survival in laryngeal cancer patients: a population-based study. Sci Rep. 2021;11:723.