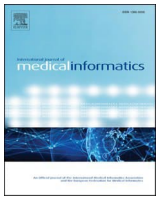
International Journal of Medical Informatics 199 (2025) 105873

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf



# Machine learning explainability for survival outcome in head and neck squamous cell carcinoma

Rasheed Omobolaji Alabi [a,b,*,1], Antti A. Mäkitie [a,c,d,1], Mohammed Elmusrati [b], Alhadi Almangush [a,e,f,g], Ylva Tiblom Ehrsson [h,1], Göran Laurell [h,1]

[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 Otorhinolaryngology – Head and Neck Surgery, University of Helsinki and Helsinki University Hospital, Helsinki, Finland
[d] Division of Ear, Nose and Throat Diseases, Department of Clinical Sciences, Intervention and Technology, Karolinska Institute and Karolinska University Hospital, Stockholm, Sweden
[e] University of Turku, Institute of Biomedicine, Pathology, Turku, Finland
[f] Department of Pathology, University of Helsinki, Helsinki, Finland
[g] Faculty of Dentistry, Misurata University, Misurata, Libya
[h] Department of Surgical Sciences, Section of Otorhinolaryngology and Head and Neck Surgery, Uppsala University, Uppsala, Sweden

ARTICLE INFO

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

ABSTRACT

*Background:* Diagnosis and treatment of head and neck squamous cell carcinoma (HNSCC) induces psychological variables and treatment-related toxicity in patients. The evaluation of outcomes is warranted for effective treatment planning and improved disease management. **Objectives**: This study aimed to build a prognostic system by combining clinicopathological parameters, treatment-related factors, and sociodemographic factors as integrative inputs to build a machine learning (ML) model to estimate the overall survival (OS) of patients with HNSCC. Furthermore, we explored the complementary prognostic potentials of these input parameters. We provide explainability and interpretability using Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) techniques. **Methods:** A total of 419 patients with HNSCC were recruited from three University Hospitals in Sweden. We compared the performance of TabNet, a state-of-the-art deep learning algorithm for tabular data, with extreme gradient boosting (XGBoost) and voting ensemble to predict OS in patients with HNSCC. **Results:** Both TabNet and XGBoost showed comparable performance accuracies, with TabNet and XGBoost showing a performance accuracy of 88.1% each and voting ensemble showing an accuracy of 88.7%. The aggregate feature importance showed that p16 (a tumor suppressor protein that plays a crucial role in cell cycle regulation), cancer stage, hemoglobin, age at diagnosis, T class, N class, smoking pack-years, body mass index (BMI), treatment modality, erythrocyte count, and human papillomavirus (HPV) status were the most important parameters for the predictive ability of the model for OS. Furthermore, we found survival trends in this cohort by individually considering parameters such as p16, cancer stage, hemoglobin, age at diagnosis, HPV status, Tumor Nodal Metastasis staging, and socioeconomic factors (marital status, housing, and level of education). In addition, both the LIME and SHAP techniques showed the contribution of each feature to the prediction made by the model. **Conclusions:** The clinical implementation of an ML model can lead to individualized risk-based therapeutic decision-making. Therefore, validating these models with multi-institutional datasets and testing them in the context of clinical trials is warranted for safe clinical implementation.

## 1. Introduction

Head and neck cancer (HNC) is a group of malignancies of the oral cavity, pharynx, larynx, salivary glands, paranasal sinuses, and nasal cavity. Most of these cancers are histologically classified as squamous cell carcinomas, collectively referred to as head and neck squamous cell

* Corresponding author at: Research Program in Systems Oncology, Faculty of Medicine, University of Helsinki, Helsinki, Finland.
*E-mail address:* rasheed.alabi@helsinki.fi (R.O. Alabi).
[1] These authors jointly supervised this work and contributed equally.

https://doi.org/10.1016/j.ijmedinf.2025.105873
Received 21 November 2024; Received in revised form 10 February 2025; Accepted 7 March 2025
Available online 8 March 2025

carcinoma (HNSCC) [1]. Patients with HNSCC often present at an advanced stage at diagnosis, which increases their mortality rate [1]. HNSCC is associated with a decreased quality of life, psychosocial concerns, disfiguration, functional impairments, treatment-related morbidity and toxicity, and mortality [2]. Besides cancer-associated causes, mortality may be related to treatment or comorbidity [3]. Hence, accurate staging and prediction of factors influencing overall survival (OS) are important for insightful treatment planning and effective management of this patient population [4].

Staging and treatment planning based on the American Joint Committee on Cancer (AJCC) tumor-node-metastasis (TNM) staging system can be used as an objective and accurate tool to predict the prognosis of patients with HNSCC. However, for individual patients, it is ineffective in predicting outcomes because of the inability of the TNM staging system to consider several important tumor- and patient-related risk factors [5]. Therefore, a comprehensive tool that integrates multiple factors is required to predict patient outcomes. In recent years, classical machine learning (ML) approaches have shown promising results in the prognostication of various cancers. In addition, deep learning (DL), a modification of classical ML, has shown potential for diagnosis, prognosis, and risk estimation in cancer management [6].

In recent years, many studies have explored the application of ML for outcome prediction. For example, the study by Yu et al., using demographic, clinical features and host factors and their model achieved a C-index of 72.9% in predicting two-year survival in patients with HNSCC [7]. Similarly, a study by Li et al., developed a ML model to predict 5-year OS of laryngeal squamous cell carcinoma (LSCC) with 85.0% accuracy [8]. In advanced LSCC, the study by Zhang et al., found that random forest ML model showed more promising OS predicting performance than the Cox model [9]. Our previous studies have also considered clinical features and demographic parameters for OS predictions in LSCC and nasopharyngeal squamous cell carcinoma using ML techniques [10,11]. In another study by Ahn et al., ML model was used to stratify oropharyngeal squamous cell carcinoma patients (OPSCC) into risk of prolonged radiation treatment duration ($\geq$ 50 days) with a promising accuracy [12]. However, most of these studies have primarily considered clinicopathologic, demographic, and treatment related factors. The present study extends previous research by incorporating pretreatment clinicopathological, laboratory, hematological, and tumor parameters, treatment-related factors, and sociodemographic factors using state-of-the-art ML algorithms. Additionally, it explores how our sociodemographic parameters affect the OS of HNSCC patients and provides explainability for a single prediction made by the model.

In this study, we used attentive interpretable tabular learning (TabNet) to estimate the OS of patients with HNSCC. We chose TabNet because of its high performance, interpretable nature for tabular data, and enhanced performance in terms of accuracy and efficiency [13,14]. The objectives of this study were to: (i)explore the complimentary prognostic potentials of combining clinicopathologic parameters, treatment-related factors alongside sociodemographic factors, as integrative input sources into TabNet, for estimating the OS of patients with HNSCC, (ii) compare the performance of TabNet with that of the extreme gradient boosting (XGBoost) and voting ensemble algorithms, which are state-of-the-art algorithms for tabular data to predict OS outcomes in patients with HNSCC (iii) provide an explanation and interpretation of the predictions made by the XGBoost model using the Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) techniques, and (iv) explore the effect of some of the input variables on OS of patients with HNSCC. The resulting explainable and interpretable model may aid in prognostication by assisting in the personalized survival stratification of patients. The overall aim was to provide an adequate individualized treatment approach for this patient population.

## 2. Materials and methods

### 2.1. Dataset

We analyzed a dataset of 419 patients with HNSCC, including those with oral, oropharyngeal, laryngeal, or hypopharyngeal cancers, obtained from a multicenter prospective observational study in Sweden. The inclusion criteria for the study were patients with newly diagnosed HNSCC who were scheduled for treatment with curative intent, aged 18 years or older, and with a performance status of 0–2, as defined by the Eastern Cooperative Oncology Group (ECOG) or World Health Organization (WHO) [15]. The exclusion criteria included a history of treatment for malignant neoplasms within the last 5 years (except for skin cancer), an inability to understand the Swedish language, severe alcohol abuse, or any cognitive impairment that might affect study compliance. Participants were included before the start of treatment between October 2015 and December 2022. Distribution of patients with HNSCC is shown in Table 1 (Sub-section 2.3).

### 2.2. Selection of patient attributes

The clinicopathological and sociodemographic variables included the Union of International Cancer Control (UICC) TNM 8th edition staging scheme (T class, N class, and M class), age at diagnosis, sex, tumor site, p16 (a tumor suppressor protein that plays a crucial role in cell cycle regulation), human papillomavirus (HPV) status, hemoglobin level, leukocyte count, erythrocyte count, percentage body weight change (loss or gain) six months before diagnosis compared to weight at diagnosis, body mass index (BMI), pack-years smoking parameter, housing, level of education, marital status, and working status (Table 1). Treatment-related parameters included surgery, radiotherapy (RT), chemoradiotherapy (CRT), combined-modality treatment, and other medical treatments (Table 1). The survival period (months from the date of diagnosis) and OS status of the patients were also recorded. From the parameters presented in Table 1, OS was considered the output of interest in this study. The output variables were relatively balanced. A detailed description of each of the included variables and categorizations is provided in Table 1. Ethical approval was obtained from the Regional Ethical Review Board of Uppsala (No. 2014/447). All the methods were performed in accordance with the 1964 Declaration of Helsinki and its subsequent amendments.

### 2.3. Data description for each variable

The study cohort for the ML model development included 419 patients with HNSCC. The baseline clinicopathological and sociodemographic characteristics obtained at the time of diagnosis are presented in Table 1. There were 295 (70.4%) males and 124 (29.6%) females with a male-to-female ratio of 2.4:1. The mean age at diagnosis was 63.4 years (SD, $\pm$ 10.8; range, 31 – 90) and the median age was 64.0 years. Regarding the treatment modalities, 60 (14.3%) patients underwent surgery, 128 (30.5%) RT, 101 (24.1%) CRT, 113 (27.0%) combined modality treatment, and 17 (4.1%) received medical treatment (cetuximab) and RT. Of the 419 patients with HNSCC, 344 (82.1%) completed the 2-year follow-up and were alive, 57 (13.6%) died before the 2-year follow-up, and 18 (4.3%) did not reach the 2-year follow-up. Altogether, 47 (11.2%) patients died of cancer and 10 (2.4%) died of other causes.

### 2.4. ML model development

The entire schematic representation for the ML model development and implementation processes presented in this study is given in Fig. 1. The process involves selection of variables for model development and using the selected variables to develop ML models based on TabNet, extreme gradient boosting (XGBoost), and voting ensemble algorithms, respectively. We compared the performance metrics of these ML models

**Table 1**
Baseline demographic characteristics of head and neck squamous cell carcinoma patients in the training data (N = 419).

| Variables | Total (N = 419) (%) | Categorization for machine learning analysis | Data type after categorization |
|---|---|---|---|
| **Age at diagnosis** | | | |
| < 40 years old (Young) | 12 (2.9) | No categorization | Integer |
| >= 40 years old (Old) | 407 (97.1) | | |
| **Sex** | | | |
| Male | 295 (70.4) | 1 = Male | Integer |
| Female | 124 (29.6) | 2 = Female | |
| **Marital status** | | | |
| Married/living together | 313 (74.7) | 1 = Married | Integer |
| Partner not living together | 18 (4.3) | 2 = Partner | |
| Single | 88 (21.0) | 3 = Single | |
| **Housing** | | | |
| House (Owning and living in a house) | 282 (67.3) | 1 = House | Integer |
| Residence flat (Condominium/owning and living in a flat | 50 (11.9) | 2 = Residence | |
| Tenancy flat (Rental apartment/renting and living in a flat) | 87 (20.8) | 3 = Tenancy | |
| **Education level** | | | |
| Lower education | 283 (67.5) | 1 = Low education | Integer |
| Higher education | 136 (32.5) | 2 = Higher education | |
| **Working situation** | | | |
| Working | 54 (12.9) | 1 | Integer |
| Pensioner (Retired) | 204 (48.7) | 2 | |
| Sick leave | 159 (37.9) | 3 | |
| Unemployed | 02 (0.5) | 4 | |
| **Pack years** | | | |
| < 20 | 298 (71.1) | No categorization | Decimal |
| >= 20 | 121 (28.9) | | |
| **Subsite/Location** | | | |
| Oropharynx | 209 (49.9) | 1 | Integer |
| Oral cavity | 147 (35.1) | 2 | |
| Larynx | 53 (12.7) | 3 | |
| Hypopharynx | 10 (12.4) | 4 | |
| **T class (UICC 8)** | | | |
| T1 | 125 (29.8) | T1 = 1 | Integer |
| T2 | 141 (33.7) | T2 = 2 | |
| T3 | 62 (14.8) | T3 = 3 | |
| T4 | 91 (21.7) | T4 = 4 | |
| **N class (UICC 8)** | | | |
| N0 | 167 (39.9) | N0 = 0 | Integer |
| N1 | 152 (36.3) | N1 = 1 | |
| N2 | 79 (18.9) | N2 = 2 | |
| N3 | 21 (5.0) | N3 = 3 | |
| **M class (UICC 8)** | | | |
| M0 | 418 (99.8) | M0 = 0 | Integer |
| M1 | 1 (0.2) | M1 = 1 | |
| **Cancer stage** | | | |
| Stage I | 166 (39.6) | Stage I = 1 | Integer |
| Stage II | 96 (22.9) | Stage II = 2 | |
| Stage III | 77 (18.4) | Stage III = 3 | |
| Stage IV | 80 (19.1) | Stage IV = 4 | |
| **P16 (Cyclin-dependent kinase inhibitor 2A, CDKN2A, multiple tumor suppressor 1)** | | | |
| Positive | 208 (49.6) | 1 = Positive | Integer |
| Negative | 96 (22.9) | 2 = Negative | |
| Non tested | 115 (27.4) | 3 = Non tested | |
| **HPV (Human Papillomavirus)** | | | |
| Positive | 117 (27.9) | 1 = Positive | Integer |
| Negative | 26 (6.2) | 2 = Negative | |
| Non tested | 276 (65.9) | 3 = Non tested | |
| **% Weight changes**§ | | | |
| Weight gain | 115 (27.5) | 1 = Positive | Integer |
| Weight loss | 304 (72.5) | 0 = Negative | |
| **BMI (Body Mass Index)** | | | |
| Severe thinness (< 16) | 0 (0.0) | No categorization | Decimal |
| Moderate thinness (16 – 17) | 2 (0.5) | | |
| Mild thinness (17.1 – 18.5) | 8 (1.9) | | |
| Normal (18.6 – 25) | 150 (35.8) | | |
| Overweight (25.1 – 30) | 167 (39.9) | | |
| Obese (> 30) | 92 (22.0) | | |
| **Leucocyte*** | Range (2.6 – 25.9) | No categorization | Decimal |
| **Erythrocyte*** | Range (0.4 – 6.1) | | Decimal |
| **Hemoglobin*** | Range (96 – 176) | | Integer |
| **Treatment modalities** | | | |
| Surgery | 60 (14.3) | 1 | Integer |
| Radiotherapy (RT) | 128 (30.5) | 2 | |
| Chemoradiotherapy (RCT) | 101 (24.1) | 3 | |
| Other medical treatment (Cetuximab or Erbitux) and RT | 17 (4.1) | 4 | |
| Combined modality treatment | 113 (27.0) | 5 | |
| **Survival** | | | |
| Cancer-related death | 47 (11.2) | 1 | Integer |
| Other | 10 (2.4) | 1 | |
| Alive | 362 (86.4) | 0 | |

* The values of leucocyte, erythrocyte, and hemoglobin were used directly without categorization for the model training.

§ Percentage body weight change (loss or gain) six months before diagnosis and at diagnosis.

(Table 2). The performance of these models to predict OS were compared based on accuracy. Furthermore, we provided explainability to the prediction made by XGBoost model using the Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) techniques. The detailed description of these processes is presented in sub-sections 2.4.1 – 2.4.3.

### 2.4.1. Variables for ML model training

The model development process is shown in the Supplementary Figure S1. The dataset for training the model consisted of (a) pretreatment parameters – age, sex, percentage change in body weight during 6 months before and at diagnosis, BMI, and pack years of smoking (b) laboratory and hematological markers – hemoglobin level, erythrocyte count, leucocyte count, (c) tumor parameters – tumor site, cancer stage and UICC TNM 8th edition staging system, p16, HPV status, and (d) sociodemographic factors – marital status, level of education, working status, and housing. The outcome of interest was OS.

### 2.4.2. TabNet for OS prediction

This study proposed TabNet, an end-to-end DL model, to predict OS in patients with HNSCC. A typical TabNet is composed of encoder and decoder blocks [14]. The encoder consists of a feature transformer, an attentive transformer, and feature-masking blocks [13,14]. The TabNet encoder involves multiple steps. First, the original input parameters (defined in **section 2.2**) were subjected to batch normalization. These normalized features were then passed through the feature transformer block. The feature transformer block is composed of a 4-layer network where two are shared across all decision steps, whereas two are decision step-dependent. Each layer consisted of a fully connected layer, batch normalization, and gated linear unit nonlinearity. Therefore, following batch normalization, the input features enter the feature transformer, where a soft selection of salient features is performed through sparse selection to make the model parameter efficient (feature masking). An attentive transformer was used to obtain the masks using the features processed in the preceding step. Ideally, an attentive transformer is composed of a fully connected layer, followed by a batch normalization layer, prior scales layer, and Sparsemax layer. While the prior scale layer contains information about the extent to which each feature has been used previously (in the current decision step), Sparsemax normalization encourages sparsity. Accuracy was used as the main metric to evaluate the performance of the model. Other important metrics are listed in Table 2.

### 2.4.3. Xgboost and voting ensemble for OS prediction

XGBoost and voting ensemble are based on ensemble learning paradigm, also known as multiple classification systems, have the potential to produce a learner that is generalizable [16]. Following the processing of data and selection of variables (sub-sections 2.1 – 2.3), the same data used for TabNet model development (sub-section 2.4.2) were used for XGBoost and voting ensemble model training. The training process was performed using 5-fold cross-validation due to the amount of dataset. This approach minimizes bias and imitates external validation [17]. Cross-validation was chosen to assess the predictive ability of the model on new data that were not used in model development, thereby helping to detect overfitting or selection bias [18]. In addition, it gives insight into how the model will generalize to an independent dataset. Technically, an ensemble paradigm ensures that multiple versions of the same machine learning model (weak or strong) are trained in such a way that each ensemble member is different (i.e., the decision trees are fit on different subsamples of the training dataset) [19]. Then, this process is followed by a selective combination of member classifiers into a better classification using any of the several appropriate and efficient ensemble methods, such as voting, averaging, bagging, stacking boosting, or boosting [16,20,21]. The XGBoost method used in this study used a boosting methodology where many moderately accurate weak learners are integrated (boosted) to form strong learning. Voting ensemble, on the other hand, is also known as a *meta-model* or *model of models* because it combines the prediction from multiple models [22]. In this study, we have used soft voting methodology, whereby it sums the predicted probabilities for each class label. The predicted class label with the largest sum probability is given as the final prediction. Hence, it is known as a majority voting ensemble [16]. Hyperparameter tuning was done where necessary to ensure that a reasonable model accuracy was achieved.

**Table 2**
Performance metrics of TabNet, XGBoost, and voting ensemble algorithms for overall survival prediction in patients with HNSCC.

| Metrics | TabNet | XGBoost | Voting ensemble |
|---|---|---|---|
| Specificity | 0.99 | 0.99 | 0.98 |
| Precision | 0.50 | 0.78 | 0.78 |
| Negative Predictive Value | 0.89 | 0.89 | 0.89 |
| False Positive Rate | 0.01 | 0.01 | 0.01 |
| False Discovery Rate | 0.50 | 0.22 | 0.22 |
| False Negative Rate | 0.90 | 0.75 | 0.75 |
| Accuracy (%) | 88.1 | 88.1 | 88.7 |

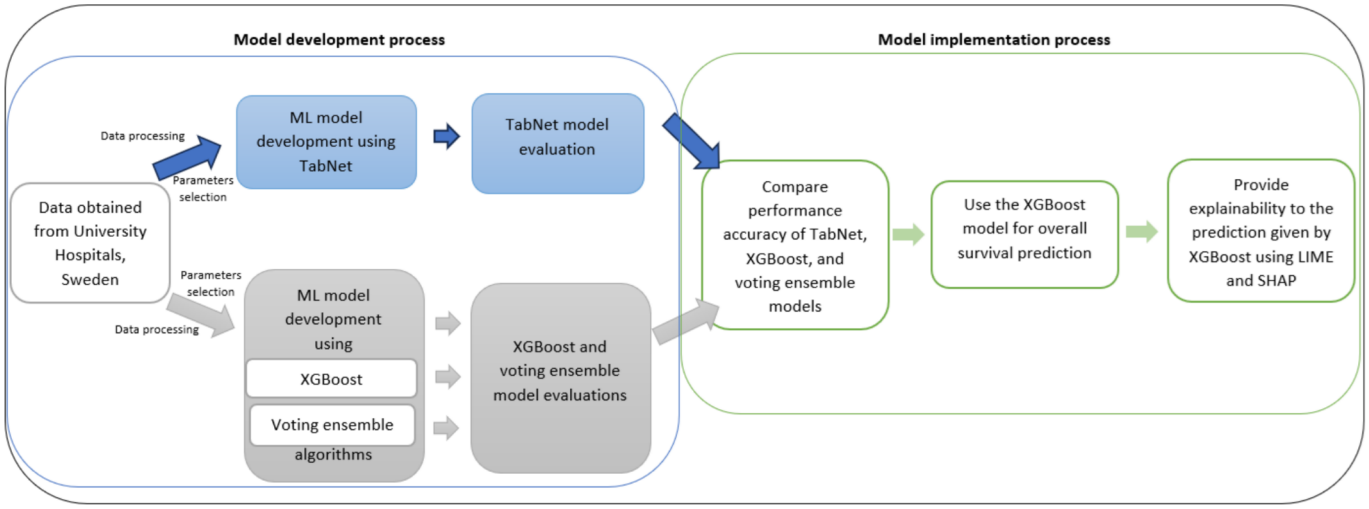



**Fig. 1.** A schematic representation of the model development.

### 2.5. Model implementation

The model implementation involves comparing the performance accuracy of ensemble machine learning models – XGBoost and voting ensemble with TabNet for predicting OS in HNSCC and providing explanations for a single prediction made by the XGBoost-based ML model. The details of model implementation are presented sub-sections 2.5.1 – 2.5.3.

#### 2.5.1. Comparison of TabNet with decision tree algorithm

In this study, we used the XGBoost algorithm as a baseline for comparison. This is due to promising results obtained from XGBoost in recent years [10,23,24]. We compared the performance of XGBoost with that of TabNet in predicting OS in patients with HNSCC. Our TabNet was implemented using PyTorch (Version 4.0, released on Sep 14, 2022) and the Scikit-learn framework.

#### 2.5.2. Local Interpretable Model-agnostic Explanations (LIME)

LIME[25], is a model-agnostic technique that is applied to an already trained model to examine the hidden relationships between the input parameters and the output parameter of interest [26]. LIME works by tweaking the input parameters while observing the effect of this tweak on the output [27]. The tweaking helps to provide local interpretability to the model prediction in terms of the degree of accuracy and the contribution of input variables to the prediction made by the model. For interpretability using LIME, we used LimeTabularExplainer in Python version 3.10.0 to fit the training data of the global model (XGBoost in this study).

#### 2.5.3. Shapley Additive exPlanations (SHAP)

SHAP uses the principle of game theory to make local explanations of the model predictions [28]. Therefore, the SHAP technique computes Shapley values by evaluating the model under several different combinations of input features and calculating the average difference in the output (prediction), when a feature is present compared to when it is absent [29]. This difference is known as the Shapley value and represents the contribution of the feature to the prediction made by the model [29]. Hence, the Shapley values quantify the contribution of each feature to the prediction of a model for a given input [27,29]. For interpretability and explainability, the SHAP technique computes the contributions of each feature to the final prediction of the decision in our XGBoost model (i.e., tree-based model). Specifically, TreeSHAP was used to estimate the Shapley values of the features in the model. These Shapley values provided a method for quantifying the contribution of each feature to the prediction made by the model (Fig. 5).

## 3. Results

### 3.1. Performance accuracy, explainability, and interpretability of XGBoost model

The performance accuracy of TabNet was 88.1%, voting ensemble was 88.7%, and XGBoost ensemble was 88.1% after model training. Other performance metrics are presented in Table 2.

### 3.2. Explainability and interpretability of XGBoost model with LIME and SHAP

In terms of explainability, for a single instance of patient survival prediction, the LIME technique showed that the model predicted patient survival with 54% probability for the correctness of the predictions made by our trained model. Furthermore, T class, cancer stage, M class, p16, HPV, treatment, erythrocyte count, and pack-years of smoking were the main parameters contributing to survival prediction made by our trained model (Fig. 2a). In contrast, for the same patient, the LIME technique showed that patient age, hemoglobin level, BMI, and change in body weight were the parameters that contributed to the model prediction of death with 46% prediction reliability (Fig. 2a).

In addition to LIME, the SHAP technique showed that erythrocyte count, treatment approach, education, marital status, and weight change during the 6 months before and at diagnosis, in order of significance had a significant effect on the model ability to predict patient survival (Fig. 2b). Similarly, the SHAP technique showed that T class, hemoglobin, cancer stage, pack-years of smoking, N class, HPV, age at diagnosis, BMI, leukocyte, p16, N class, cancer stage, hemoglobin, leukocyte count, and tumor site, in order of significance, had a lower impact on a single instance of prediction by the model (Fig. 2b). The SHAP and LIME techniques presented herein provide a local explanation (for a single instance of a patient with HNC) (Fig. 2a-b).

### 3.3. Evaluating the input variables for aggregate feature importance

The aggregate feature importance of input variables for predicting OS in patients with HNSC were p16, cancer stage, hemoglobin level, age at diagnosis, T class, N class, pack-years, BMI, treatment, erythrocyte count, and HPV status of the tumor (Fig. 3). Of note, Fig. 3 represents the most important result of this study, as it shows how clinicopathologic parameters, treatment, and sociodemographic factors contributed to the OS of patients with HNSCC.

### 3.4. Evaluating the input variables for individual feature importance

Evaluation of the importance of individual features showed that a positive p16 or non-tested p16 tumor status was associated with better survival than a negative p16 status (Fig. 4a). In addition, cancer stages I-III were associated with a better OS than stage IV (Fig. 4b). We found that the survival of patients with HNSCC decreased as hemoglobin levels increased to above 150 g per liter (g/l) (Fig. 5a). Moreover, the survival of patients with HNSCC decreased steadily as their age increased. However, a sharp increase was observed between 65 and 70 years of age (Fig. 5b). Similarly, positive and non-tested HPV tumor statuses were associated with better survival (Fig. 6). Regarding the TNM staging scheme, patients with T1-3 had better survival rates than those with T4 (Fig. 7a). Likewise, patients with N1 had better survival than those with N0, whereas both N0 and N1 were associated with better OS than N2 (Fig. 7b). Patients with smoking pack-years of 10–20 had better survival than those with $>$ 20 pack-years of smoking (Fig. 8a). Regarding treatment modalities, patients with HNC who received surgery, radiotherapy, or other treatments plus radiotherapy had better OS than those who received chemotherapy alone or in combination (Fig. 8b). Considering the survival of patients with HNC in terms of socioeconomic parameters, those who were married or living together and those with a partner but not living together had better survival rates than those who were single (Supplementary Figure S2a). In addition, patients owning or living in a house or tenancy flat had better survival rates than those living in a flat residence (Supplementary Figure 2b). No significant survival trends were observed in patients with HNSCC with respect to the following parameters: sex, level of education, working status, site, M class, percentage body weight change (percentage weight gain or weight loss during 6 months before and at diagnosis), BMI, leukocyte count, and erythrocyte count.

## 4. Discussion

Individualized therapeutic decision-making based on survival prognosis remains challenging in the management of patients with HNSCC. This is due to the complex and heterogeneous nature of the disease itself as well as patient- and treatment-related factors. Therefore, in this study, we included only patients with HNSCC to ensure that we focused on a histologically homogeneous group of patients. Our model examined the complex relationships between these parameters by analyzing how similar patients with histologically confirmed squamous cell carcinoma

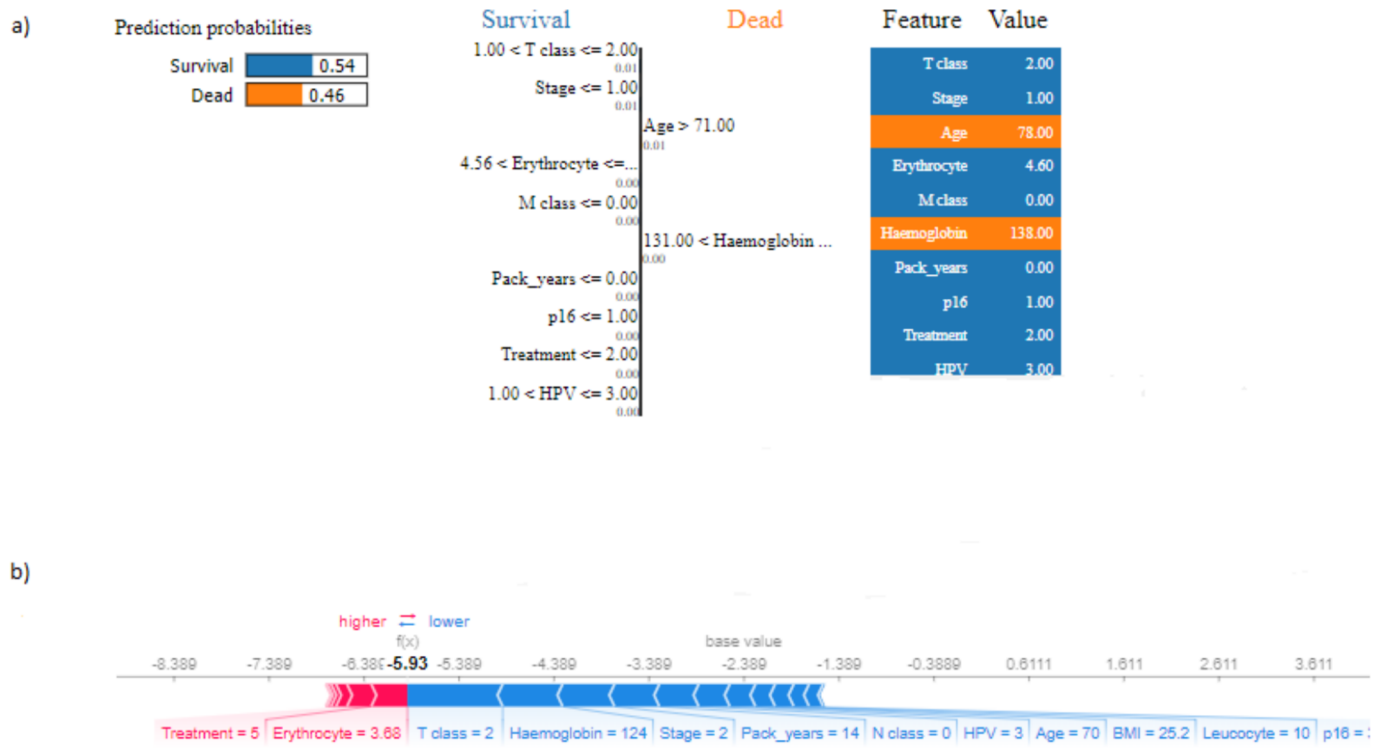



**Fig. 2.** LIME and SHAP frameworks for explainability and interpretability.

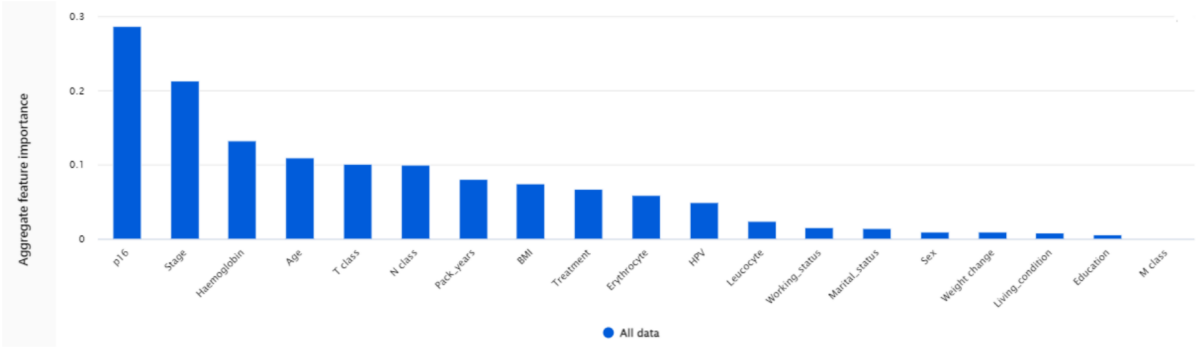



**Fig. 3.** Aggregate feature importance of input variables.

responded in the prospective cohort, thereby predicting the outcome of the new patients under consideration. Therefore, the results fulfilled our ultimate goal of aiding personalized treatment plans for patients with HNSCC.

This study used TabNet, a state-of-the-art DL model for tabular data, to enhance the prediction of OS in patients with HNSCC. The model was developed using the TabNet algorithm and demonstrated a promising performance. The performance of TabNet is comparable to the traditional voting ensemble method, which has been reported to be suitable for a variety of classification tasks [30]. We explored the LIME and SHAP techniques to analyze how the input variables contributed to the survival prediction made by the model. This is important for personalized prognostication of the survival outcomes in patients with HNSCC.

Multiple studies have found that p16 is an important prognostic surrogate biomarker in oropharyngeal HNSCC [31–33]. This supports the results of our study, in which the developed ML model identified p16 as the most important parameter. This may be because the cohort used to develop the ML model included a significant number of patients with oropharyngeal cancer. Notably, 73% of patients were tested for p16. We found that patients with p16-positive tumors had better survival rates than those with p16-negative tumors. Our findings are consistent with previous studies that emphasized HPV-related status as an important prognostic and predictive biomarker, mainly in oropharyngeal squamous cell carcinoma (OPSCC), potentially enabling better prognostication and choice of therapeutic options [34]. We found that patients with HPV-positive tumors had better survival rates. Over the years, a significant number of HNSCC cases have been associated with chronic alcohol or tobacco consumption. However, in recent years, an increasing proportion of OPSCC cases has been caused by mucosal infections with high-risk HPV [34]. This presents a challenge for the optimization of clinical management. In general, patients with HPV-related diseases are younger (often non-consumers of either alcohol or tobacco) and respond better to different treatment options, resulting in a better overall prognosis [34].

In case of non-oropharyngeal HNSCC, several conflicting results have been reported regarding the prognostic significance of p16 [35]. Several studies have reported that p16 has prognostic significance [21,24–28], whereas other studies have reported contrasting results [36–39]. Due to these contrasting views, practice guidelines generally recommend routine testing for p16 in oropharyngeal HNSCC, but not in non-

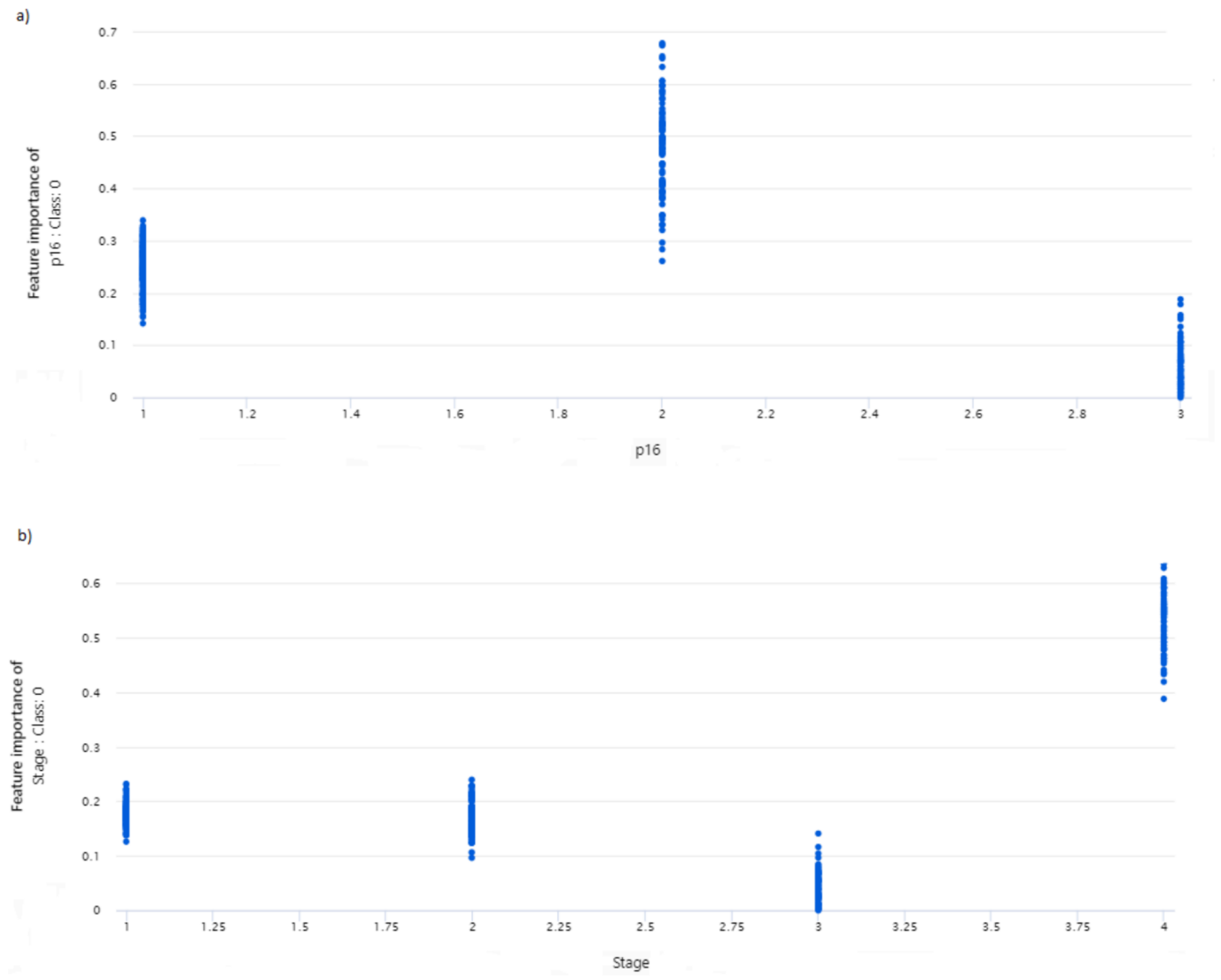



**Fig. 4.** Individual feature importance for (a) p16 and (b) cancer stage of HNSCC.

oropharyngeal HNSCC [35].

Our ML model highlighted the stage of HNSCC at presentation as the second most important determinant of outcomes. This finding has been corroborated by several studies emphasizing the clinical stage at presentation as a major predictor of survival in HNSCC [40]. A significant number of patients with HNSCC are diagnosed at an advanced stage of cancer with an increased risk of recurrence and distant metastasis [1,40–42]. Therefore, a predictive model can aid in early intervention and treatment planning. Having prior knowledge of the survival outcomes of a patient can assist clinicians in personalized treatment planning and enhance insightful decision-making [10]. As shown in Fig. 4b, patients with stage I-III cancer had better survival rates. Stage III cancer was associated with improved survival, which may be attributed to the aggressive treatments that are usually administered to patients with advanced-stage HNC.

Treatment- and tumor-related factors may lead to psychological disorders that may negatively impact appetite and, consequently, result in poor nutritional status. Remarkably, this has been associated with tumor progression and poor survival in these patients [43]. Patients with poor nutritional value are likely to have reduced weight and a low BMI. Ottosson et al. reported that low and normal BMI were negative factors for OS in patients with OPSCC [44]. In the case of OPSCC, moderate or severe weight loss at diagnosis has been reported as a prognostic factor for recurrence [45]. Therefore, nutritional interventions are warranted in patients with HNSCC [43,46]. Interestingly, nutritional and hematological factors have been found to have prognostic value in patients with HNSCC receiving curative treatment [43]. Lim et al. found that body weight, serum total protein and albumin levels, and hematological variables significantly affected OS [43]. This finding was justified in the present study, in which the developed model identified hematological markers such as hemoglobin levels, erythrocyte count, and percentage body weight change in HNC as important prognostic markers for OS. Further studies are warranted to evaluate malnutrition and weight loss during diagnosis and treatment.

Our results suggest that hemoglobin level at diagnosis plays a significant role in the OS of patients with HNSCC. Umesh et al. reported that low baseline hemoglobin ($< 12.7$ g/dl) is a poor predictive marker for relapse-free survival and OS in patients with HNSCC (treated with definitive radiotherapy, with or without chemotherapy) [47]. Monitoring and correcting hemoglobin levels may offer an accessible means to improve treatment outcomes at a low cost and effort and should thus be considered in future studies [47,48]. Hemoglobin level is thought to serve as a systemic inflammation marker that may predict treatment outcomes and guide nutritional intervention in these patients [49].

Furthermore, our model predicted that patient age at diagnosis is an important prognostic factor for OS. This corresponds to the findings of

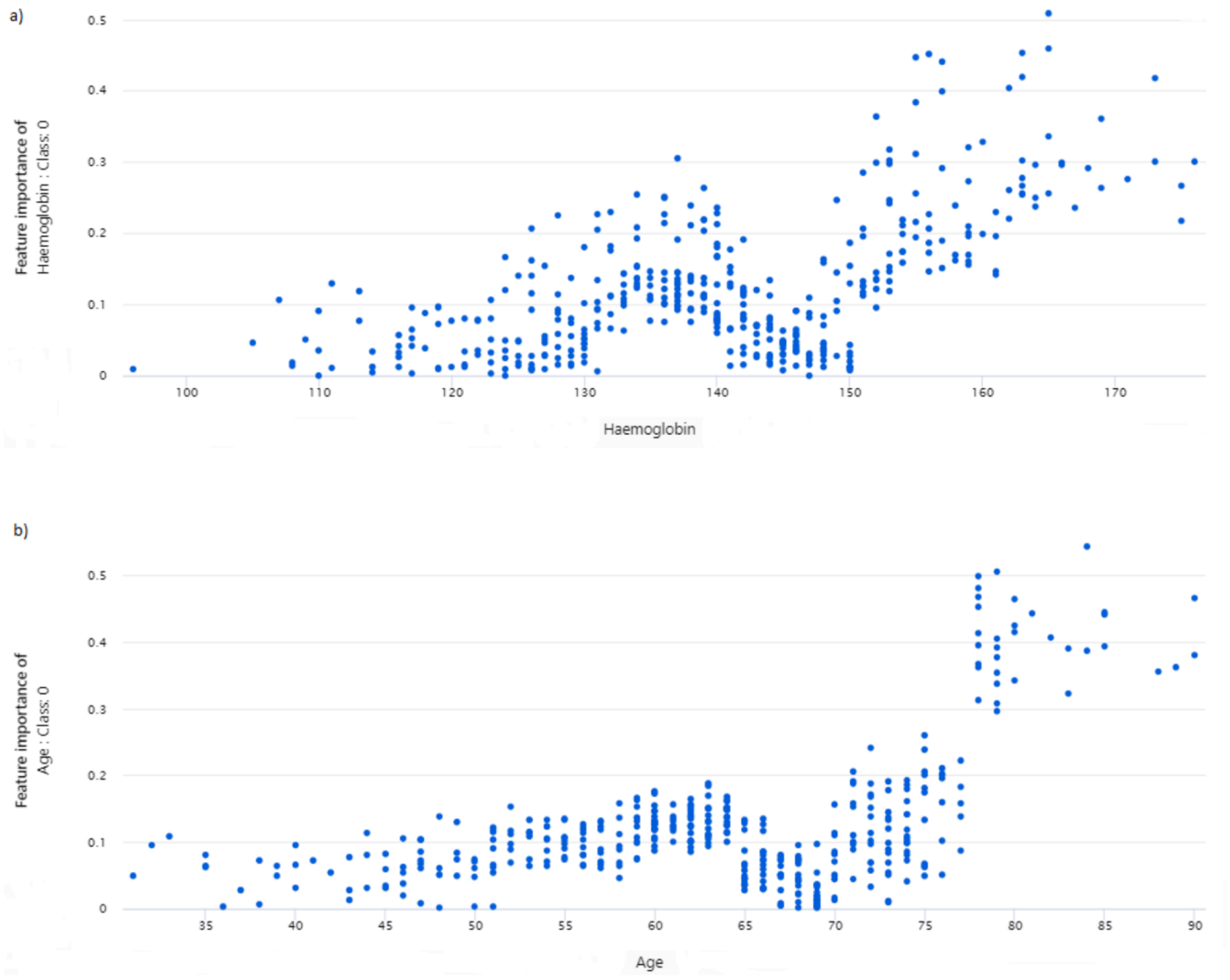



**Fig. 5.** Individual feature importance for (a) hemoglobin level and (b) age at diagnosis.

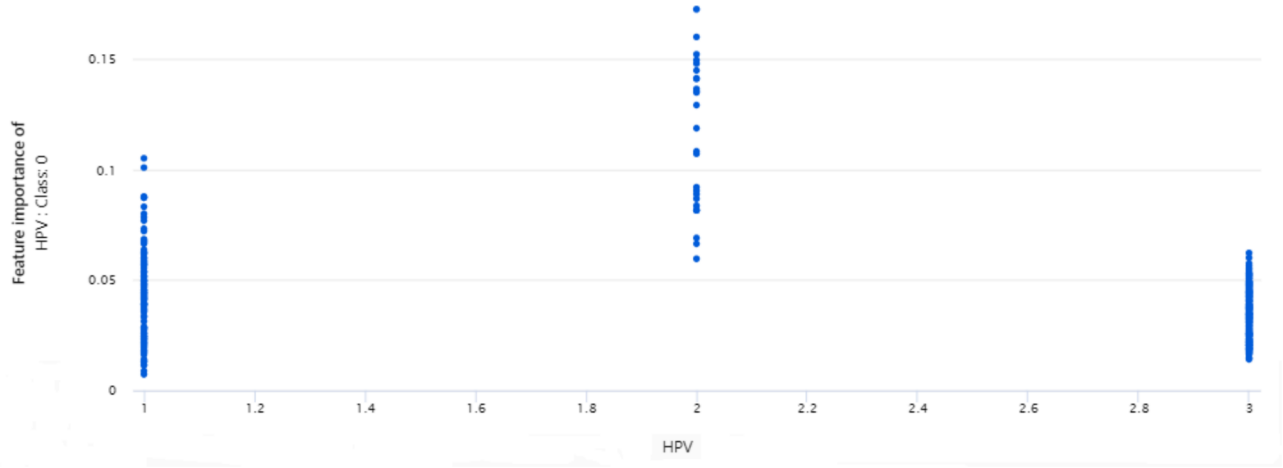



**Fig. 6.** Individual feature importance for HPV.

Cadoni et al., in which increasing age was found to be an important factor for OS in HNC [41]. Similarly, Talani et al. showed that age is a risk factor for early death among patients with HNSCC with curative treatment intent [50]. We observed a reduction in OS as the age of the patients with HNC increased. However, there was a sudden increase in the OS of patients aged–65–70 years. This may have been due to the more aggressive treatment administered to this cohort. Although older patients may experience worse treatment-related toxicity, especially in

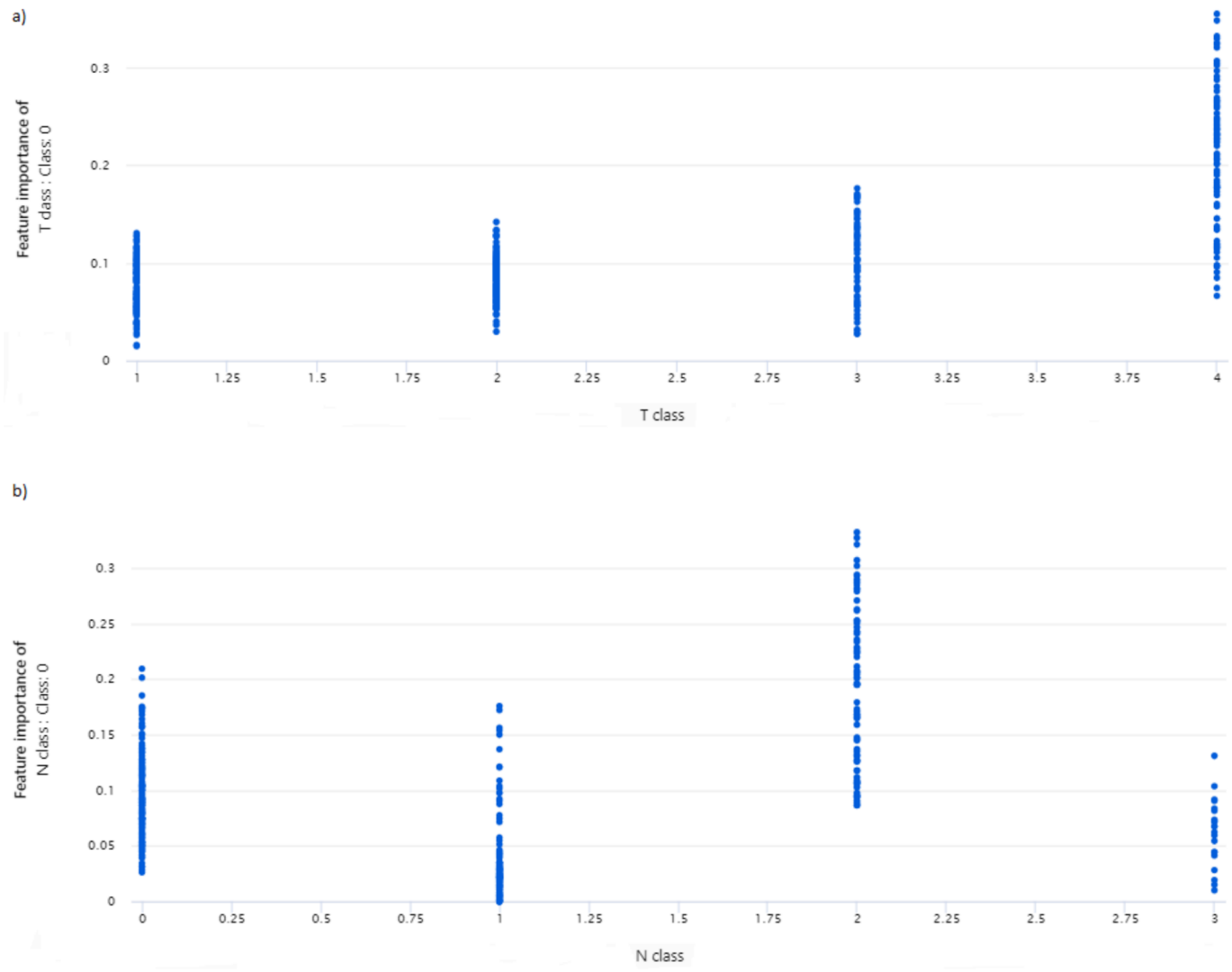



**Fig. 7.** Individual feature importance for (a) T class and (b) N class.

terms of treatment intensification, some studies have reported that older patients have equal survival outcomes compared with their younger peers [41]. Therefore, there is a growing debate suggesting that chronological age alone should not withhold curative treatment intention in most patients with HNSCC [51]. This debate is buoyed by the paucity of randomized data regarding the effect of age on treatment response, treatment-related morbidity, and OS in patients with HNC [41]. Additionally, a recent study by Lop et al. reported that disease-specific survival starts to decrease significantly in patients with HNSCC after the age of 80 [51,52].

As indicated by our model, the tumor and nodal classes are important prognostic factors. We used the 8th edition of the UICC TNM staging system. T classes 1–2 showed improved survival compared to T classes 3–4. Notably, some patients with class T3 tumors had better survival. This may be due to the disparity in classification and staging; some cases of T2 tumors may be misstaged and should instead be classified as T3 and receive an enhanced treatment approach. Additionally, the incorporation of depth of invasion into the T class later caused a shift to upstaging [53]. Similarly, patients with classes N0-1 showed improved survival. Some patients with the N1 class had better survival rates than those with N0 class. This may be associated with the integration of extranodal extensions (ENE) into N class [53]. Therefore, the UICC TNM 8 remains an objective and accurate classification scheme because it already incorporates the depth of invasion into T staging and ENE into N staging [53]. The inclusion of ENE in N staging is expected to reflect the burden of neck nodal disease accurately. In addition, the UICC 8th edition was found to be more suitable for HPV-associated carcinoma [54].

The present model identified the number of pack-years smoked as an important prognostic factor associated with estimating survival in patients with HNSCC. This finding is consistent with earlier reports showing that tobacco smoking reduces the efficacy of radiation therapy within HNC cohorts [55,56]. Given the significance of the pack-year threshold in determining patients eligible for treatment deintensification [55], several conflicting reports have emerged regarding smoking exposure. This may largely depend on the HNC sites. For example, Ma et al. reported 22 pack-years as the threshold for estimating cancer treatment outcomes [55]. In other reports, approximately 10 pack-years was found to be a risk factor for early disease recurrence in patients with oral cancer [57].

In our study, it is also notable that socioeconomic status (SES), such as housing, working status, level of education, marital status, and sex were not among the top (one to ten) variables identified by our model as being associated with OS, which is inconsistent with other studies; socioeconomic factors and demographic disparities may be significant as predictors of survival in patients with HNC [58–62]. This discrepancy in the effect of SES as a predictor of survival in HNC may be due to a high standard of living, minimal social and economic disparities, availability

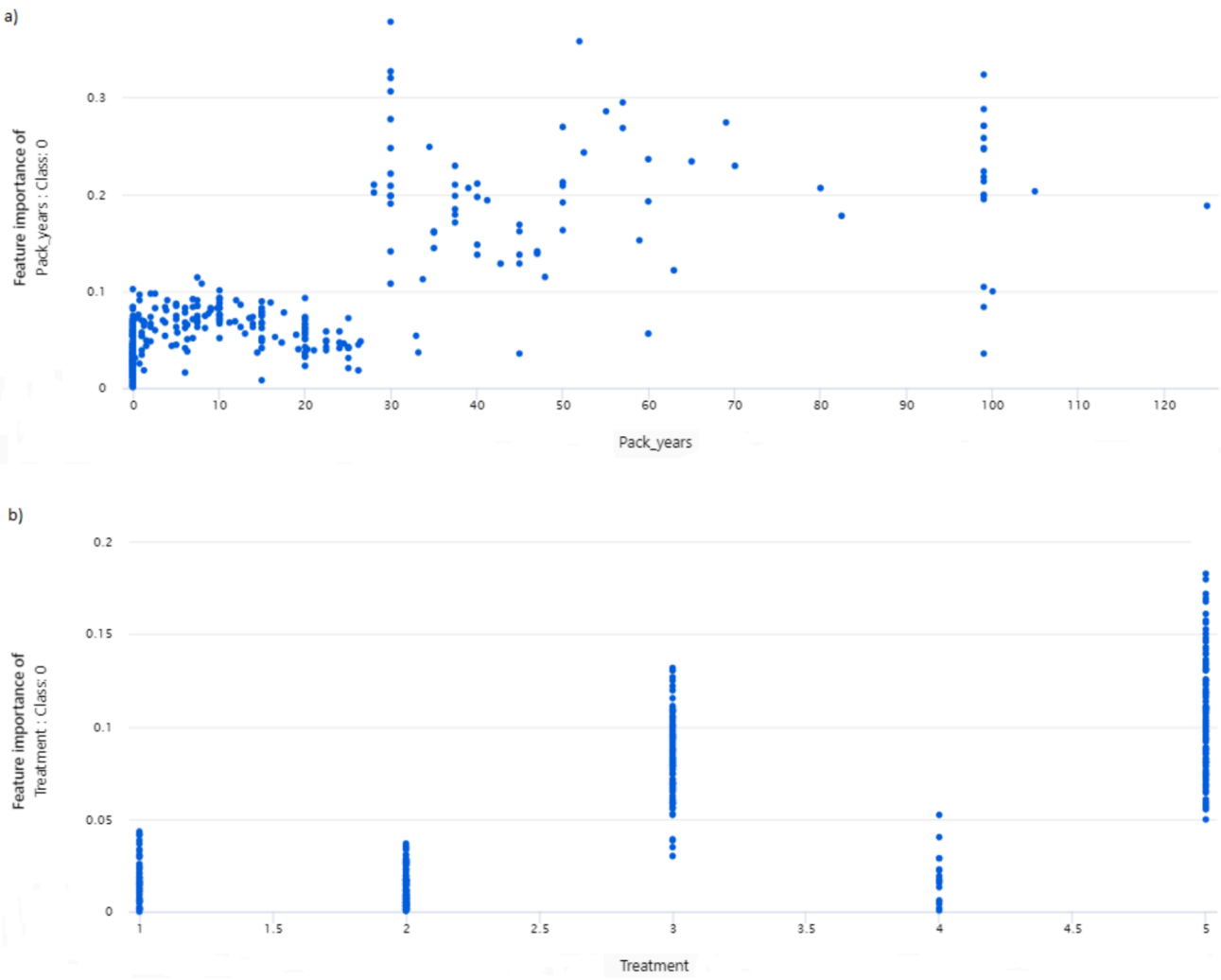



**Fig. 8.** Individual feature importance for (a) pack years and (b) treatment modalities.

of moderately sufficient resources for adequate cancer surveillance, high health literacy, access to care, uniform treatment modality, and low health disparity, which are widely prevalent in the majority of the Nordic countries where our data are from. Therefore, the relationship between SES and poor OS is complex and multifactorial [59]. Future research is warranted to validate and elucidate the mechanism and relationship between SES and poor survival to identify possibilities for evidence-based interventions.

There are several limitations in the implementation of ML in daily clinical practice that warrant further investigation [63,64]. Our study further emphasized the potential of artificial intelligence and ML for outcome prognostication and personalized medicine to improve HNC management. One of the major and continued criticisms is that the obtained predictions are often nontransparent and uninterpretable. Hence, we incorporated LIME and SHAP to explain and interpret the predictions of the XGBoost model. The model predicts patient survival with 54% accuracy. Furthermore, the model provided interpretability and explainability by showing that T and M classes, cancer stage, p16, and erythrocyte count contributed to predicting patient survival with 54% accuracy. Conversely, the age of the patient, BMI, hemoglobin level, and body weight change were considered by the model to suggest (at 46% accuracy) that the patient may die. Similarly, we provided explainability and interpretability for another patient using SHAP. The erythrocyte count, treatment approach, education, marital status, and weight change during the 6 months before and at diagnosis (in decreasing order of significance) had a greater effect on the model's ability to predict patient survival. In contrast, T class, hemoglobin, cancer stage, pack-years of smoking, N class, HPV, age at diagnosis, BMI, leukocyte count, p16, and tumor site (in decreasing order of significance) had a lower impact on the model's prediction for that single instance of prediction using SHAP. Ultimately, based on these analyses of model explainability with LIME and SHAP, the clinician/oncologist is able to consider the two arguments (interpretability and explainability) made by the model and to make a final therapeutic clinical decision.

In this study, we compared a DL approach (TabNet) for tabular data with two traditional ensemble ML algorithms that have been previously applied to numerous tasks involving tabular data. TabNet consists of a 4-layer network that includes a feature transformer, attentive transformer, and masking. Because TabNet is based on feature selection, it performs on par with the traditional ensemble ML models such as voting ensemble and XGBoost, which is based on portioning samples for model training. With a relatively large number of datasets, the effect of TabNet components can be significant. Hence, TabNet may outperform traditional ensemble methods.

Our study has some limitations. First, as expected, the number of cases used for the ML model development was not uniform among HNSCC sites and the outcome of interest, which may lead to a relatively high false negative rate. In future, the model may be retrained using a more balanced dataset of tumor subsites and OS outcome. Second, further external geographic validation using a relatively large number of cases is warranted to explore the true performance of the model.

In conclusion, our model showed promising performance in

predicting the OS of patients with HNSCC. This model may serve as an accurate, objective, and ancillary tool at a low cost for clinicians, for enhanced decision-making [65]. Our model incorporated multiple clinicopathological, treatment-related, and socioeconomic parameters to estimate patient survival. The model developed using these combined parameters may address disparities in the currently used cancer staging schemes and aid in choosing a personalized management approach. Furthermore, the explainability and interpretability provided by the LIME and SHAP frameworks may enhance opinions of clinicians to better understand how these parameters contribute to decisions or predictions. Such a strategy is important for individualized treatment planning in patients with HNSCC. In future studies, the developed model may be integrated as a web-based tool for testing purposes by other research centers to encourage research collaboration. In addition, model performance can be improved through federated learning in the future.

## 5. Summary points

- We developed a model for predicting overall survival (OS) in patients with head and neck squamous cell carcinoma (HNSCC) by combining clinicopathological parameters, treatment-related factors, and sociodemographic factors.
- We compared two ensemble ML algorithms – extreme gradient boosting (XGBoost) and voting ensemble algorithms with TabNet for OS prediction in patients with HNSCC patients.
- Explainability and interpretability of the model were enhanced using the Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive explanation (SHAP) techniques.
- Aggregate feature importance showed that p16, cancer stage, hemoglobin level, age at diagnosis, T class, N class, pack-years of smoking, BMI, treatment, erythrocyte count, and HPV were the most important parameters.
- Clinical implementation of an ML model can lead to individualized risk-based therapeutic decision making.

## CRediT authorship contribution statement

**Rasheed Omobolaji Alabi:** Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Antti A. Mäkitie:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Mohammed Elmusrati:** Writing – review & editing. **Alhadi Almangush:** Writing – review & editing, Methodology. **Ylva Tiblom Ehrsson:** Writing – review & editing, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Göran Laurell:** Writing – review & editing, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgment

The study was supported by the Sigrid Jusélius Foundation (AM 240138), Finska Läkaresällskapet (AM 2024), The Finnish State Research Funding to the Helsinki University Hospital (TYH2024203), and The Swedish Cancer Society (grant numbers 2015/363, 2018/502, and 21 1419 Pj).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijmedinf.2025.105873.

## Data availability

The datasets generated in this study are available from the corresponding author upon reasonable request.

## References

[1] D.E. Johnson, B. Burtness, C.R. Leemans, V.W.Y. Lui, J.E. Bauman, J.R. Grandis, Head and neck squamous cell carcinoma, Nat Rev Dis Primers 6 (2020) 92, https://doi.org/10.1038/s41572-020-00224-3.
[2] A.A. Mäkitie, R.O. Alabi, L. Pulkki-Råback, A. Almangush, J.J. Beitler, N.F. Saba, P. Strojan, R. Takes, O. Guntinas-Lichius, A. Ferlito, Psychological factors related to treatment outcomes in head and neck cancer, Adv Ther (2024), https://doi.org/10.1007/s12325-024-02945-3.
[3] J. Lop, M.D.P. Venegas, A. Pujol, B. Sauter, R. Vásquez, M. Casasayas, M. Quer, X. León, Causes of long-term mortality in patients with head and neck squamous cell carcinomas, Eur Arch Otorhinolaryngol 279 (2022) 3657–3664, https://doi.org/10.1007/s00405-021-07211-8.
[4] R.O. Alabi, M. Elmusrati, I. Leivo, A. Almangush, A.A. Mäkitie, Artificial intelligence-driven radiomics in head and neck cancer: current status and future prospects, Int. J. Med. Inf. 188 (2024) 105464, https://doi.org/10.1016/j.ijmedinf.2024.105464.
[5] R.O. Alabi, A.A. Mäkitie, M. Pirinen, M. Elmusrati, I. Leivo, A. Almangush, Comparison of nomogram with machine learning techniques for prediction of overall survival in patients with tongue cancer, Int. J. Med. Inf. 145 (2021) 104313, https://doi.org/10.1016/j.ijmedinf.2020.104313.
[6] R.O. Alabi, A. Almangush, M. Elmusrati, A.A. Mäkitie, Deep machine learning for oral cancer: from precise diagnosis to precision medicine, Front. Oral. Health 2 (2022) 794248, https://doi.org/10.3389/froh.2021.794248.
[7] H. Yu, S.J. Ma, M. Farrugia, A.J. Iovoli, K.E. Wooten, V. Gupta, R.P. McSpadden, M. A. Kuriakose, M.R. Markiewicz, J.M. Chan, W.L. Hicks, M.E. Platek, A.K. Singh, Machine learning incorporating host factors for predicting survival in head and neck squamous cell carcinoma patients, Cancers 13 (2021) 4559, https://doi.org/10.3390/cancers13184559.
[8] Z. Li, T. Li, P. Zhang, X. Wang, A practical online prediction platform to predict the survival status of laryngeal squamous cell carcinoma after 5 years, Am. J. Otolaryngol. 45 (2024) 104209, https://doi.org/10.1016/j.amjoto.2023.104209.
[9] Y.-F. Zhang, Y.-J. Shen, Q. Huang, C.-P. Wu, L. Zhou, H.-L. Ren, Predicting survival of advanced laryngeal squamous cell carcinoma: comparison of machine learning models and Cox regression models, Sci Rep 13 (2023) 18498, https://doi.org/10.1038/s41598-023-45831-8.
[10] R.O. Alabi, M. Elmusrati, I. Leivo, A. Almangush, A.A. Mäkitie, Machine learning explainability in nasopharyngeal cancer survival using LIME and SHAP, Sci Rep 13 (2023) 8984, https://doi.org/10.1038/s41598-023-35795-0.
[11] R.O. Alabi, A. Almangush, M. Elmusrati, I. Leivo, A.A. Mäkitie, Interpretable machine learning model for prediction of overall survival in laryngeal cancer, Acta Otolaryngol. (2024) 1–7, https://doi.org/10.1080/00016489.2023.2301648.
[12] S. Ahn, E.J. Oh, M.I. Saleem, T. Tham, Machine learning methods in classification of prolonged radiation therapy in oropharyngeal cancer: national cancer database, Otolaryngol.–Head Neck Surg. 171 (2024) 1764–1772, https://doi.org/10.1002/ohn.926.
[13] R. Shahriarirad, S.M. Meshkati Yazd, R. Fathian, M. Fallahi, Z. Ghadiani, N. Nafissi, Prediction of sentinel lymph node metastasis in breast cancer patients based on preoperative features: a deep machine learning approach, Sci Rep 14 (2024), https://doi.org/10.1038/s41598-024-51244-y.
[14] S.Ö. Arik, T. Pfister, TabNet: attentive interpretable tabular learning, AAAI 35 (2021) 6679–6687, https://doi.org/10.1609/aaai.v35i8.16826.
[15] M.M. Oken, R.H. Creech, D.C. Tormey, J. Horton, T.E. Davis, E.T. McFadden, P. P. Carbone, Toxicity and response criteria of the Eastern Cooperative Oncology Group, Am J Clin Oncol 5 (1982) 649–655.
[16] V.C. Osamor, A.F. Okezie, Enhancing the weighted voting ensemble algorithm for tuberculosis predictive diagnosis, Sci Rep 11 (2021) 14806, https://doi.org/10.1038/s41598-021-94347-6.
[17] Y.-J. Tseng, H.-Y. Wang, T.-W. Lin, J.-J. Lu, C.-H. Hsieh, C.-T. Liao, Development of a machine learning model for survival risk stratification of patients with advanced oral cancer, JAMA Netw. Open 3 (2020) e2011768, https://doi.org/10.1001/jamanetworkopen.2020.11768.
[18] G. Cawley, N. Talbot, On over-fitting in model selection and subsequent selection bias in performance evaluation, J. Mach. Learn. Res. 11 (2010) 2079–2107.
[19] J. Brownlee, A gentle introduction to ensemble learning, (2020). https://machinelearningmastery.com/what-is-ensemble-learning/ (accessed August 19, 2022).
[20] F. Aydın, Z. Aslan, The construction of a majority-voting ensemble based on the interrelation and amount of information of features, Comput. J. 63 (2020) 1756–1774, https://doi.org/10.1093/comjnl/bxz118.
[21] Y. Zhang, H. Zhang, J. Cai, B. Yang, A weighted voting classifier based on differential evolution, Abstract and Applied Analysis 2014 (2014) 1–6, https://doi.org/10.1155/2014/376950.
[22] J. Brownlee, How to develop voting ensembles with Python, (2020). https://machinelearningmastery.com/voting-ensembles-with-python/ (accessed August 19, 2022).

[23] T. Chen, C. Guestrin, XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, San Francisco California USA, 2016: pp. 785–794. DOI: 10.1145/2939672.2939785.
[24] R.O. Alabi, A. Sjöblom, T. Carpén, M. Elmusrati, I. Leivo, A. Almangush, A. A. Mäkitie, Application of artificial intelligence for overall survival risk stratification in oropharyngeal carcinoma: a validation of ProgTOOL, Int. J. Med. Inf. 175 (2023) 105064, https://doi.org/10.1016/j.ijmedinf.2023.105064.
[25] M.T. Ribeiro, S. Singh, C. Guestrin, "Why Should I Trust You?": Explaining the Predictions of Any Classifier, arXiv:1602.04938 [Cs, Stat] (2016). http://arxiv.org/abs/1602.04938 (accessed July 25, 2021).
[26] A. Gramegna, P. Giudici, SHAP and LIME: an evaluation of discriminative power in credit risk, Front. Artif. Intell. 4 (2021) 752558, https://doi.org/10.3389/frai.2021.752558.
[27] R. Alabi, A. Almangush, M. Elmusrati, I. Leivo, A.A. Mäkitie, An interpretable machine learning prognostic system for risk stratification in oropharyngeal cancer, Int. J. Med. Inf. 168 (2022) 104896, https://doi.org/10.1016/j.ijmedinf.2022.104896.
[28] S. Lundberg, S.-I. Lee, A Unified Approach to Interpreting Model Predictions, arXiv: 1705.07874 [Cs, Stat] (2017). http://arxiv.org/abs/1705.07874 (accessed July 25, 2021).
[29] A. Joseph, Parametric inference with universal function approximators, (2019). DOI: 10.48550/ARXIV.1903.04209.
[30] K. Kita, T. Fujimori, Y. Suzuki, Y. Kanie, S. Takenaka, T. Kaito, T. Taki, Y. Ukon, M. Furuya, H. Saiwai, N. Nakajima, T. Sugiura, H. Ishiguro, T. Kamatani, H. Tsukazaki, Y. Sakai, H. Takami, D. Tateiwa, K. Hashimoto, T. Wataya, D. Nishigaki, J. Sato, M. Hoshiyama, N. Tomiyama, S. Okada, S. Kido, Bimodal artificial intelligence using TabNet for differentiating spinal cord tumors: integration of patient background information and images, iScience 26 (2023), https://doi.org/10.1016/j.isci.2023.107900.
[31] C. Ndiaye, M. Mena, L. Alemany, M. Arbyn, X. Castellsagué, L. Laporte, F.X. Bosch, S. De Sanjosé, H. Trottier, HPV DNA, E6/E7 mRNA, and p16INK4a detection in head and neck cancers: a systematic review and meta-analysis, Lancet Oncol. 15 (2014) 1319–1331, https://doi.org/10.1016/S1470-2045(14)70471-1.
[32] B. O'Sullivan, S.H. Huang, J. Su, A.S. Garden, E.M. Sturgis, K. Dahlstrom, N. Lee, N. Riaz, X. Pei, S.A. Koyfman, D. Adelstein, B.B. Burkey, J. Friborg, C.A. Kristensen, A.B. Gothelf, F. Hoebers, B. Kremer, E.-J. Speel, D.W. Bowles, D. Raben, S. D. Karam, E. Yu, W. Xu, Development and validation of a staging system for HPV-related oropharyngeal cancer by the International Collaboration on Oropharyngeal cancer Network for Staging (ICON-S): a multicentre cohort study, Lancet Oncol. 17 (2016) 440–451, https://doi.org/10.1016/S1470-2045(15)00560-4.
[33] G. D'Souza, W.H. Westra, S.J. Wang, A. Van Zante, A. Wentz, N. Kluz, E. Rettig, W. R. Ryan, P.K. Ha, H. Kang, J. Bishop, H. Quon, A.P. Kiess, J.D. Richmon, D. W. Eisele, C. Fakhry, Differences in the prevalence of human papillomavirus (HPV) in head and neck squamous cell cancers by sex, Race, Anatomic Tumor Site, and HPV Detection Method, JAMA Oncol 3 (2017) 169, https://doi.org/10.1001/jamaoncol.2016.3067.
[34] J.P. Kühn, W. Schmid, S. Körner, F. Bochen, S. Wemmert, H. Rimbach, S. Smola, J. C. Radosa, M. Wagner, L.G.T. Morris, V. Bozzato, A. Bozzato, B. Schick, M. Linxweiler, HPV status as prognostic biomarker in head and neck cancer-which method fits the best for outcome prediction? Cancers (Basel) 13 (2021) 4730, https://doi.org/10.3390/cancers13184730.
[35] A.K. Bryant, E.J. Sojourner, L.K. Vitzthum, K. Zakeri, H. Shen, C. Nguyen, J. D. Murphy, J.A. Califano, E.E.W. Cohen, L.K. Mell, Prognostic role of p16 in nonoropharyngeal head and neck cancer, JNCI: Journal of the National Cancer Institute 110 (2018) 1393–1399, https://doi.org/10.1093/jnci/djy072.
[36] C.H. Chung, Q. Zhang, C.S. Kong, J. Harris, E.J. Fertig, P.M. Harari, D. Wang, K. P. Redmond, G. Shenouda, A. Trotti, D. Raben, M.L. Gillison, R.C. Jordan, Q.-T. Le, p16 protein expression and human papillomavirus status as prognostic biomarkers of nonoropharyngeal head and neck squamous cell carcinoma, JCO 32 (2014) 3930–3938, https://doi.org/10.1200/JCO.2013.54.5228.
[37] L. Sivars, C. Bersani, N. Grün, T. Ramqvist, E. Munck-Wikland, C. Von Buchwald, T. Dalianis, Human papillomavirus is a favourable prognostic factor in cancer of unknown primary in the head and neck region and in hypopharyngeal cancer, Molecular and Clinical Oncology 5 (2016) 671–674, https://doi.org/10.3892/mco.2016.1050.
[38] L. Satgunaseelan, S.A. Virk, T. Lum, K. Gao, J.R. Clark, R. Gupta, p16 expression independent of human papillomavirus is associated with lower stage and longer disease-free survival in oral cavity squamous cell carcinoma, Pathology 48 (2016) 441–448, https://doi.org/10.1016/j.pathol.2016.03.015.
[39] A.J. Feinstein, S.G. Shay, E. Chang, M.S. Lewis, M.B. Wang, Treatment outcomes in veterans with HPV-positive head and neck cancer, Am. J. Otolaryngol. 38 (2017) 188–192, https://doi.org/10.1016/j.amjoto.2017.01.005.
[40] A.A. Adoga, D.D. Kokong, N.D. Ma'an, J.G. Mugu, C.J. Mgbachi, A.M. Dauda, The predictive factors of primary head and neck cancer stage at presentation and survival in a developing nation's tertiary hospital, Sage Open Med. 6 (2018) 205031211879241, https://doi.org/10.1177/2050312118792416.
[41] G. Cadoni, L. Giraldi, L. Petrelli, M. Pandolfini, M. Giuliani, G. Paludetti, R. Pastorino, E. Leoncini, D. Arzani, G. Almadori, S. Boccia, Prognostic factors in head and neck cancer: a 10-year retrospective analysis in a single-institution in Italy, Acta Otorhinolaryngol Ital 37 (2017) 458–466, https://doi.org/10.14639/0392-100X-1246.
[42] C.C. Reyes-Gibby, K.O. Anderson, K.W. Merriman, K.H. Todd, S.S. Shete, E. Y. Hanna, Survival patterns in squamous cell carcinoma of the head and neck: pain as an independent prognostic factor for survival, J. Pain 15 (2014) 1015–1022, https://doi.org/10.1016/j.jpain.2014.07.003.
[43] W.S. Lim, J. Roh, S. Kim, S. Choi, S.Y. Nam, S.Y. Kim, Pretreatment albumin level predicts survival in head and neck squamous cell carcinoma, Laryngoscope 127 (2017), https://doi.org/10.1002/lary.26691.
[44] S. Ottosson, K. Söderström, E. Kjellén, P. Nilsson, B. Zackrisson, G. Laurell, Weight and body mass index in relation to irradiated volume and to overall survival in patients with oropharyngeal cancer: a retrospective cohort study, Radiat Oncol 9 (2014) 160, https://doi.org/10.1186/1748-717X-9-160.
[45] A. Baptistella, K. Hilleshein, C. Beal, J. Brambatti, R. Caron, S. Baptistella, R. Zuquello, C. Rossoni, G. Manfro, Weight loss as a prognostic factor for recurrence and survival in oropharyngeal squamous cell carcinoma patients, Mol Clin Onc (2018), https://doi.org/10.3892/mco.2018.1737.
[46] A.A. Mäkitie, R.O. Alabi, H. Orell, O. Youssef, A. Almangush, A. Homma, R. P. Takes, F. López, R. de Bree, J.P. Rodrigo, A. Ferlito, Managing cachexia in head and neck cancer: a systematic scoping review, Adv Ther 39 (2022) 1502–1523, https://doi.org/10.1007/s12325-022-02074-9.
[47] P. Umesh, S. Tandon, S. Purohit, P. Ahlawat, 181 Baseline Hemoglobin as an independent marker for survival outcome in squamous cell carcinoma of head and neck- An institutional review from tertiary cancer centre, Radiother. Oncol. 192 (2024) S213–S216, https://doi.org/10.1016/S0167-8140(24)00532-2.
[48] C. Melo-Alvim, P. Miguel-Semedo, R.S. Paiva, S. Lobo-Martins, H. Luna-Pais, A. L. Costa, A.R. Santos, A. Florindo, A.L. Vasconcelos, A.N. Abrunhosa-Branquinho, P. Palmela, L. Fernandes, D.L. Presa, L. Costa, L. Ribeiro, Pretreatment hemoglobin level as a prognostic factor in patients with locally advanced head and neck squamous cell carcinoma, Reports of Practical Oncology & Radiotherapy 25 (2020) 768–774, https://doi.org/10.1016/j.rpor.2020.07.002.
[49] A. Rühle, E. Haehl, H. David, T. Kalckreuth, T. Sprave, R. Stoian, C. Zamboglou, E. Gkika, A. Knopf, A.-L. Grosu, N.H. Nicolay, The value of laboratory parameters for anemia, renal function, systemic inflammation and nutritional status as predictors for outcome in elderly patients with head-and-neck cancers, Cancers 12 (2020) 1698, https://doi.org/10.3390/cancers12061698.
[50] C. Talani, A. Mäkitie, M. Beran, E. Holmberg, G. Laurell, L. Farnebo, Early mortality after diagnosis of cancer of the head and neck – A population-based nationwide study, PLoS One 14 (2019), https://doi.org/10.1371/journal.pone.0223154.
[51] J. Lop, C. Valero, J. García, M. Quer, I. Ganly, J.P. Shah, S.G. Patel, X. León, Does age influence disease-specific survival in patients with squamous cell carcinomas of the head and neck? J. Surg. Oncol. 121 (2020) 1058–1066, https://doi.org/10.1002/jso.25895.
[52] P.R. Camilon, W.A. Stokes, S.A. Nguyen, E.J. Lentsch, The prognostic significance of age in oropharyngeal squamous cell carcinoma, Oral Oncol. 50 (2014) 431–436, https://doi.org/10.1016/j.oraloncology.2013.12.013.
[53] A.Y. Pai, A. Rizvi, D.G. Bryant, K. Mitsimponas, UICC TNM8 implementation and its effects on head and neck cancer staging in a tertiary referral centre, Advances in Oral and Maxillofacial Surgery 1 (2021) 100022, https://doi.org/10.1016/j.adoms.2021.100022.
[54] N. Würdemann, S. Wagner, S.J. Sharma, E.-S. Prigge, M. Reuschenbach, S. Gattenlöhner, J.P. Klussmann, C. Wittekindt, Prognostic impact of AJCC/UICC 8th edition new staging rules in oropharyngeal squamous cell carcinoma, Front Oncol 7 (2017) 129, https://doi.org/10.3389/fonc.2017.00129.
[55] S.J. Ma, H. Yu, B. Yu, O. Waldman, M. Khan, U. Chatterjee, S. Santhosh, J. Gill, A. J. Iovoli, M. Farrugia, A. Shevorykin, E. Carl, K. Wooten, V. Gupta, R. McSpadden, M.A. Kuriakose, M.R. Markiewicz, A. Al-Afif, W.L. Hicks, M.E. Platek, M. Seshadri, C. Sheffer, G.W. Warren, A.K. Singh, Association of pack-years of cigarette smoking with survival and tumor progression among patients treated with chemoradiation for head and neck cancer, JAMA Netw Open 5 (2022) e2245818, https://doi.org/10.1001/jamanetworkopen.2022.45818.
[56] A.J. Platek, V. Jayaprakash, M. Merzianu, M.E. Platek, D.M. Cohan, W.L. Hicks, S. P. Marimuthu, T.B. Winslow, V. Gupta, H. Arshad, M.A. Kuriakose, S. Dibaj, J. R. Marshall, M.E. Reid, G.W. Warren, A.K. Singh, Smoking cessation is associated with improved survival in oropharynx cancer treated by chemoradiation, Laryngoscope 126 (2016) 2733–2738, https://doi.org/10.1002/lary.26083.
[57] K. Sambasivan, I. Sassoon, S. Thavaraj, R. Kennedy, G. Doss, A. Michaelidou, E. Odell, A. Sandison, G. Hall, P. Morgan, L.H.C. Collins, A. Lyons, L. Cascarini, A. Fry, R. Oakley, R. Simo, J.-P. Jeannon, I. Petkar, M. Reis Ferreira, A. Kong, M. Lei, T. Guerrero Urbano, TNM 8 staging is a better prognosticator than TNM 7 for patients with locally advanced oral cavity squamous cell carcinoma treated with surgery and post-operative radiotherapy, Radiother. Oncol. 160 (2021) 54–60, https://doi.org/10.1016/j.radonc.2021.04.003.
[58] S.H. Choi, J.E. Terrell, K.E. Fowler, S.A. McLean, T. Ghanem, G.T. Wolf, C. R. Bradford, J. Taylor, S.A. Duffy, Socioeconomic and other demographic disparities predicting survival among head and neck cancer patients, PLoS One 11 (2016), https://doi.org/10.1371/journal.pone.0149886.
[59] N.R. Lenze, J.T. Bensen, W.G. Yarbrough, A.G. Shuman, Characteristics and outcomes associated with anxiety and depression in a head and neck cancer survivorship cohort, Am. J. Otolaryngol. 43 (2022) 103442, https://doi.org/10.1016/j.amjoto.2022.103442.
[60] B. Weizman, N. Golan, O. Ronen, Effect of socioeconomic status on survival in patients with head and neck cancer, Head Neck 43 (2021) 3001–3009, https://doi.org/10.1002/hed.26788.
[61] U.C. Megwalu, Impact of county-level socioeconomic status on oropharyngeal cancer survival in the United States, Otolaryngol.–Head Neck Surg. 156 (2017) 665–670, https://doi.org/10.1177/0194599817691462.
[62] J.M. Rotsides, J.R. Oliver, L.E. Moses, M. Tam, Z. Li, D. Schreiber, A.S. Jacobson, K. S. Hu, B. Givi, Socioeconomic and racial disparities and survival of human papillomavirus–associated oropharyngeal squamous cell carcinoma, Otolaryngol.–

Head Neck Surg. 164 (2021) 131–138, https://doi.org/10.1177/0194599820935853.
[63] R.O. Alabi, O. Youssef, M. Pirinen, M. Elmusrati, A.A. Mäkitie, I. Leivo, A. Almangush, Machine learning in oral squamous cell carcinoma: Current status, clinical concerns and prospects for future—A systematic review, Artif. Intell. Med. 115 (2021) 102060, https://doi.org/10.1016/j.artmed.2021.102060.
[64] A.A. Mäkitie, R.O. Alabi, S.P. Ng, R.P. Takes, K.T. Robbins, O. Ronen, A.R. Shaha, P.J. Bradley, N.F. Saba, S. Nuyts, A. Triantafyllou, C. Piazza, A. Rinaldo, A. Ferlito, Artificial intelligence in head and neck cancer: a systematic review of systematic reviews, Adv Ther 40 (2023) 3360–3380, https://doi.org/10.1007/s12325-023-02527-9.
[65] M. Akcay, D. Etiz, O. Celik, A. Ozen, Evaluation of prognosis in nasopharyngeal cancer using machine learning, Technol. Cancer Res. Treat. 19 (2020), https://doi.org/10.1177/1533033820909829.