Original Article

Using a machine learning algorithm and clinical data to predict the risk factors of disease recurrence after adjuvant treatment of advanced-stage oral cavity cancer

Sheng-Yao Huang[a,b], Ren-Jun Hsu[a,c,d], Dai-Wei Liu[a,b,c,d], Wen-Lin Hsu[b,c,d]*

[a]Institute of Medical Sciences, Tzu Chi University, Hualien, Taiwan, [b]Department of Radiation Oncology, Hualien Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation, Hualien, Taiwan, [c]Cancer Center, Hualien Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation, Hualien, Taiwan, [d]School of Medicine, Tzu Chi University, Hualien, Taiwan

Abstract

Objectives: Head-and-neck cancer is a major cancer in Taiwan. Most patients are in the advanced stage at initial diagnosis. In addition to primary surgery, adjuvant therapy, including chemotherapy and radiotherapy, is also necessary to treat these patients. We used a machine learning tool to determine the factors that may be associated with and predict treatment outcome. **Materials and Methods:** We retrospectively reviewed 187 patients diagnosed with advanced-stage head-and-neck cancer who received surgery and adjuvant radiotherapy with or without chemotherapy. We used eXtreme Gradient Boosting (XGBoost) – a gradient tree-based model – to analyze data. The features were extracted from the entries we recorded from the electronic health-care system and paper medical record. The patient data were categorized into training and testing datasets, with labeling according to their recurrence status within the 5-year follow-up. The primary endpoint was to predict whether the patients had recurrent disease. The risk factors were identified by analyzing the feature importance in the model. For comparison, we also used regression to perform the variate analysis to identify the risk factors. **Results:** The accuracy, sensitivity, and positive predictive value of the model were 57.89%, 57.14%, and 44.44%, respectively. Pathological lymph node status was the most important feature, followed by whether the patient was receiving chemotherapy. Fraction size, early termination, and interruption were the important factors related to radiotherapy and might affect treatment outcome. The area under the curve of the receiver operating characteristic curve was 0.58. The risk factors identified by XGBoost were consistent with those found by regression. **Conclusion:** We found that several factors were associated with treatment outcome in advanced-stage head-and-neck cancer. In future, we hope to collect the data according to the features introduced in this study and to construct a stronger model to explain and predict outcomes.

Keywords: *eXtreme Gradient Boosting, Head-and-neck cancer, Machine learning*

Introduction

Head-and-neck cancer is one of the 10 cancers with the highest incidence rates and mortality rates in Taiwan [1] and is mainly caused by social habits, including smoking, chewing betel nuts, and drinking alcohol. In Taiwan, head-and-neck cancer occurs mostly in the oral cavity, oropharynx, hypopharynx, and larynx. Although there are policies for early screening, about 30% of the cases are initially diagnosed when they are at an advanced stage [1]. For these cases, surgery is the main option if medically operable, while adjuvant concurrent chemoradiotherapy (CCRT) is often essential [2], especially for those with risk factors such as positive surgical margins or extranodal extension (ENE, previously known as extracapsular spreading [ECS]) found in pathology [3-5]. However, most previous research has focused on defining subgroups suitable for receiving postoperative treatment. From the perspective of radiation oncology, finding the risks or even predicting patients who may be prone to recurrent disease after treatment is crucial.

*Address for correspondence: Dr. Wen-Lin Hsu,
Department of Radiation Oncology, Hualien Tzu Chi Hospital,
Buddhist Tzu Chi Medical Foundation, 707, Section 3, Chung-Yang
Road, Hualien, Taiwan.
E-mail: 109353113@gms.tcu.edu.tw

Access this article online

Quick Response Code:

Website: www.tcmjmed.com

DOI: 10.4103/tcmj.tcmj_56_24

Machine learning has been developed as a mature tool for evaluating data, including medical data [6]. It can be categorized into supervised and unsupervised learning. In supervised learning, humans usually provide annotations to model as the "ground truth" to learn and optimize. Unsupervised learning, on the other hand, often lacks right answers, and the model separates the data using mathematical parameters, such as distance on vector space. Machine learning generally performs several tasks depending on human needs. For example, classification involves dividing data into discrete categories. Choosing the appropriate model and task for the data is essential.

We propose a method of predicting disease recurrence using a machine learning model. We also determine the potential risks associated with recurrence by analyzing the model's feature importance. Based on our findings, we hope to gather more data in future for a more definitive and comprehensive conclusion.

MATERIALS AND METHODS

We retrospectively recruited patients diagnosed with oral cavity cancer, with the subsite including the buccal mucosa, gingivae, tongue, and hard palate, from January 1, 2005, to December 21, 2016. All patients were diagnosed at Hualien Tzu Chi General Hospital, with tissue proof (mainly biopsy), image studies, and primary surgery. Imaging modalities included computed tomography (CT, magnetic resonance imaging for evaluating local and regional extension of the disease, chest X-ray, abdominal sonography, and whole-body bone scan for distant metastasis. Primary surgery included tumor-wide resection, mandibulectomy or maxillectomy, and neck dissection. All specimens were reviewed by two pathologists to establish a pathological diagnosis. We recruited patients with pathological stage IVA and IVB disease. For the inclusion in the study, the patients should have received adjuvant radiotherapy. Table 1 shows the characteristics of the patients.

We obtained the data from each patient's medical record – both the health-care information system record and the paper medical record. We noted each patient's basic characteristics, including gender, age, cancer history, and the presence of comorbidities, and calculated the Charlson Comorbidity Index (CCI) score [7]. We also collected the information on diagnosis, including the date of clinical diagnosis and pathological diagnosis, clinical and pathological stage (including tumor-node-metastasis stage, based on the 7th edition of the American Joint Committee on Cancer [AJCC] staging manual), imaging modalities, subsites, tissue histology and differentiation, surgical method, pathological risk factors including perineural invasion (PNI), lymphovascular permeation, and ENE. The chemotherapy regimen and delivery dose were also recorded. We also obtained the data on radiotherapy, including fraction size, cumulative dose, duration of radiotherapy, days of interruption, and whether radiotherapy was terminated early. Finally, we collected the data on outcome: local, regional, and distant recurrence and the date confirmed; date of death of deceased

Table 1: Patient characteristics

Characteristics	*n*	Proportion (%)
Gender		
Male	169	90
Female	18	10
CCI score		
2	58	31
3	75	40
4	40	21
5	11	6
6	3	2
Subsite		
Tongue	57	30
Gingivae	61	33
Retromolar trigone	5	3
Hard palate	4	2
Lip	45	24
Buccal mucosa	14	7
Mouth floor	1	0.5
Clinical stage		
T1	2	1
T2	18	10
T3	18	10
T4a	134	72
T4b	14	8
N0	93	50
N1	32	17
N2	44	24
N3	17	9
Pathological stage		
T2	2	1
T3	1	0.5
T4a	177	95
T4b	7	4
N0	91	49
N1	27	15
N2	52	28
N3	16	9
Surgical method		
Mandibulectomy	103	55
Maxillectomy	34	18
Ipsilateral neck dissection	180	97
Contralateral neck dissection	46	25
Histology: Squamous cell carcinoma	186	99
Differentiation		
Well	90	48
Moderate	89	47
Poor	5	3
Pathological risk factor		
LVP	124	66
PNI	124	66
ENE	128	68
Positive margin	3	2
Chemotherapy	135	72
Radiotherapy		
Total dose <59.4 Gy	9	5
Fraction size 1.8 Gy	57	30
Fraction size 2 Gy	129	69

Contd...

Table 1: Contd...

Characteristics	*n*	Proportion (%)
Interruption >14 days	4	3
Early termination	4	3
Treatment failure within 5 years	66	35

LVP: Lymphovascular permeation, PNI: Perineural invasion,
ENE: Extranodal extension, CCI: Charlson Comorbidity Index

patients, and whether the death is directly caused by disease progression. This study was conducted in accordance with the Declaration of Helsinki and data collection was approved by the Institutional Review Board of Hualien Tzu Chi General Hospital, No. IRB106-51-B. Informed consent was waived by the IRB. The title of the protocol reviewed was "Retrospective study for postoperative radiotherapy patients of head and neck cancer," and it has been approved since May 01, 2017.

We used eXtreme Gradient Boosting (XGBoost), which is a supervised tree-based model, for the classification task [8]. The framework of XGBoost is composed of regression trees. Let f_n be functions belonging to functional space F, mapping input features x to output $f_n(x)$:

$$\hat{y} = \theta(x_i) = \sum_{n=1}^{N} f_n(x_i), f_n \in F \quad (1)$$

where ŷ indicates the prediction of the model. Thus, the prediction is decided by all the trees in the model. The prediction is compared to the label, or ground truth, to evaluate the gap. There are several ways to evaluate the gap or so-called loss function:

$$L(\theta) = \sum_i l\left(\hat{y}_i, y_i\right) + \sum_n \Omega(f_n)$$

where $\Omega(f) = \gamma T + \frac{1}{2}\lambda \|w\|^2$ \quad (2)

The loss function usually adds a term (Ω) to prevent the model from overfitting during training, and T represents the number of leaves in the tree. The procedure is called regularization. A method named additive manner is used to tackle the problem that equation (2) cannot be optimized in the traditional way. Let $\hat{y}_i^k$ be the prediction of the i-th instance in the k-th iteration, f_k is added to equation (2) to minimize the objective as follows:

$$L^{(k)} = \sum_i l\left(\hat{y}_i^{(k-1)}, y_i + f_k(x_i)\right) + \Omega(f_k)$$

We can approach the loss by the Taylor series in secondary order:

$$L^{(k)} \cong \sum_i \left(l(\hat{y}_i^{(k-1)}, y_i) + g_i f_k(x_i) + \frac{1}{2} h_i f_k^2(x_i) \right) + \Omega(f_k) \quad (3)$$

where $g_i = \partial_{\hat{y}^{(k-1)}} l(\hat{y}_i^{(k-1)}, y_i)$ and $h_i = \partial_{\hat{y}^{(k-1)}}^2 l(\hat{y}_i^{(k-1)}, y_i)$ are the first- and second-order partial derivatives of lost function. The constant term $l(\hat{y}_i^{(k-1)}, y_i)$ could be removed during approximation. Furthermore, we can rewrite Ω according to

equation (2) and make the equation as follows:

$$\tilde{L}^{(k)} = \sum_i \left(g_i f_k(x_i) + \frac{1}{2} h_i f_k^2(x_i) \right) + \gamma T + \frac{1}{2}\lambda \sum_j w_j^2$$

$$= \sum_j \left[\left(\sum_{i \in I_j} g_i \right) w_i + \frac{1}{2}\left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T \quad (4)$$

where we define I_j as the output set of leaf j of the model by the i-th instance. As a result, we can get the weight of leaf j, w_j, by

$$w_j = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (5)$$

and compute optimal value by

$$\tilde{L}^{(k)} = -\frac{1}{2} \sum_j \frac{\left(\sum_{i \in I_j} g_i\right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \quad (6)$$

The losses are compared in different leaves in a node, and the split is established by the lowest loss it can get. Subsequently, the model is constructed.

Now, there are several unique features that make XGBoost advanced to other tree-based algorithms:

1. Approximate algorithm on split finding. The conventional split finding algorithm, used by Random Forest or GBM, is an exact greedy algorithm. It iterates all possible splits by all the features and compares the score gained by each split. The split with the maximal score will be decided. Although the exact greedy algorithm gets a precise split by not missing any possible splits, the computation time and resources are large. As data volume increases, the calculation may be insufficient due to memory limitations. Thus, an approximate algorithm is proposed to tackle the problem in XGBoost. In summary, the algorithm first selects split points by the percentile of feature distribution. Then, the algorithm fills in data by chunk according to split points and finds the best solution. Approximate algorithm can be used in global (splits are proposed before tree construction) or local (splits are proposed after tree construction), each suit for different conditions. By fine-tuning the approximation factor, the algorithm has a noninferior performance when compared to the exact greedy algorithm

2. Sparsity-aware split finding. It is often that data from practical situations with missing values. In XGBoost, a default direction is introduced in each tree node. When a value is missing for the model to classify, the default direction will be applied. The model sets the direction by the data. In this way, the approximate algorithm will only visit nonmissing values and the model is not necessary to take a sparse matrix into account.

3. Column block. Among tree learning models, no matter what split algorithm is used, data will be sorted by the values before computation. Sorting data, therefore, becomes a critical step to determine the process time of the model. XGBoost uses memory units called blocks to store data. Based on the approximate algorithm, chunked data are stored in blocks, which are queued in a compressed column. Each column corresponds to a feature. The model can then define split

points by linear scan to data. In addition, the model becomes more sufficient by computing each column parallelly

4. Cache-aware access. The main idea is to prevent cache misses caused by immediate read/write dependencies. In approximate algorithms, the problem is solved by adjusting block size. An undersized block prolongs the computation and affects efficiency. An oversized block, however, may lead to an overload to memory, and cache loss occurs. The moderate size has been chosen by the XGBoost model from the multiple experiments.

Back to our study, the settings of model training and datasets were as follows. The input for the model was the data collected, with most data being converted to categorical values. The model could integrate both categorical variables and continuous variables. The output was also the primary endpoint to predict whether the patient would develop recurrent disease within the next 5 years. When building a model, we could adjust the threshold of samples to be split and limit the number of leaves or the depth of the tree to prevent the model from overfitting. Although we tried several settings of the number of samples, trees, and depth, we finally chose to let the model decide the settings by the approximate algorithm. We set the error rate as an evaluation metric, which was calculated by the number of wrong cases divided by the number of all cases. According to the sklearn manual, it was the most suitable way to evaluate model loss in binary classification. The settings of the model and training are shown in Table 2. We constructed the model and analyzed data on Python 3.7, with the package sklearn [9]. We used the package pandas to read and preprocess data. Finally, the results were illustrated using the package matplotlib.

For comparison, we also used a regression model to find out potential risk factors. We used Cox proportional hazards regression models to estimate the effects of risk factors on the hazard ratios (HRs) accompanying 95% confidence intervals. All the models were adjusted for the covariates (age at diagnosis, tumor cell differentiation, pN, margin status, lymphovascular invasion, PNI, ECS, and CT). The level of statistical significance was set at $P < 0.05$, two-tailed. The statistical method was performed on SAS 9.4.

We used accuracy, sensitivity, and positive predictive value (PPV) as evaluation metrics for the XGBoost model. Accuracy was evaluated with the following equation:

$$accuracy = \frac{a}{t},$$

where a was the number of samples correctly predicted by the model and t was the total number of samples.

Both sensitivity and PPV were derived from the confusion matrix [Table 3].

The equation for sensitivity was

$$sensitivity = \frac{TP}{TP + FN}$$

The equation for PPV was

$$PPV = \frac{TP}{TP + FP}$$

Finally, we plotted the receiver operating characteristic (ROC) curve to calculate the area under the curve (AUC).

RESULTS

We initially enrolled 206 patients, of whom 19 were excluded due to insufficient information on staging or treatment in the medical record. Finally, 187 patients were analyzed. We then selected the features from the dataset and preserved representative, independent features based on domain knowledge. As Figure 1 shows, the distribution of features did not differ between annotations in the binary classification.

Table 2: Settings of model and training

Parameters	Settings
Number of trees	Default*
Learning rate	0.01
Max tree depth	Default*
Max leaves	Default*
Subsample for trees	0.7
Evaluation metrics	Error rate
Early stopping round	50

*Default: Decided by model

Table 3: The confusion matrix

Ground truth	Model prediction	
	Positive	**Negative**
Positive	TP	FN
Negative	FP	TN

TP: True positive, NP: False positive, TN: True negative, FN: False negative
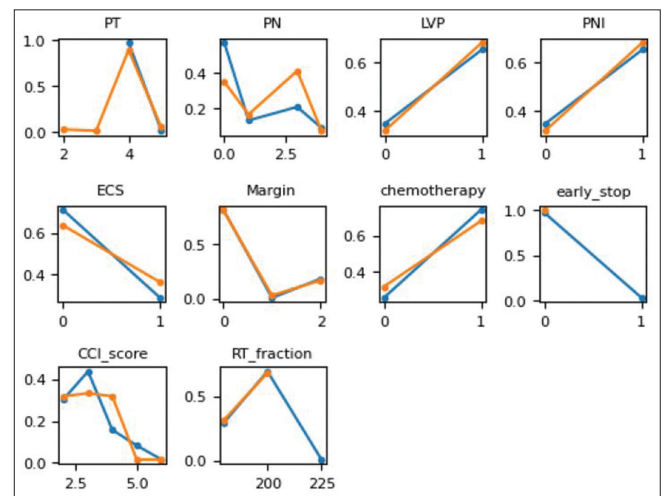

Figure 1: Feature distribution by annotation class. Blue: label = 0, no recurrence within 5 years. Orange: label = 1, recurrence within 5 years. The y-axis represents the ratio of patients in each category to patients in each dataset. PNI: Perineural invasion, ECS: Extracapsular spreading, CCI: Charlson comorbidity index, LVP: Lymphovascular permeation, PT: Pathological T stage, PN: Pathological N stage, RT: Radiotherapy

We split the data into a training dataset (168 patients) and a testing dataset (19 patients). The data in the testing dataset were not used in model training. We compared the features between the two datasets, as shown in Figure 2. Figure 3 presents a flowchart of the data processing.

The model obtained an accuracy, sensitivity, and PPV of 100% after being trained on the training dataset. We did not manually fine-tune the hyperparameters. Model inference on the testing dataset yielded an accuracy, sensitivity, and PPV of 57.89%, 57.14%, and 44.44%, respectively. We, then, analyzed the feature importance of the model [Figure 4]. The feature with the highest weighting in the model was pathological lymph node status, followed by whether the patient received chemotherapy. Fraction size, early termination, and interruption were the important factors

Figure 2: Feature distribution by training and testing dataset. Blue: Training dataset. Orange: Testing dataset. The y-axis represents the ratio of patients in each category to patients in each dataset. PNI: Perineural invasion, ECS: Extracapsular spreading, CCI: Charlson comorbidity index, LVP: Lymphovascular permeation, PT: Pathological T stage, PN: Pathological N stage, RT: Radiotherapy

related to radiotherapy. The ROC curve of the model indicated an AUC of 0.58 [Figure 5].

The risk factors defined by the regression model are shown in Table 4. Pathological lymph node status was the only factor that might affect treatment outcomes (adjusted HR 2.22, $P = 0.009$).

DISCUSSION

XGBoost is a widely used machine learning tool for classification tasks [10] and is also applied to medical data [8,11]. As a gradient-boosting model, it processes the data much faster than other models. It can deal with classification and regression data at the same time, and most importantly, it integrates several classifiers to form a model, which generates built-in ensemble learning, and to increase its robustness.

There is debate about the issues of complexity and the black box in machine learning. As a model becomes more complex, more parameters are added to increase layers (to go deeper) or divisions in a layer (to go wider). Complex models lead to two problems: the interpretability of the model and the trap of overfitting. In the case of interpretability, with increasing complexity of the calculation mechanism of the model, it may become too difficult to explain or doubt may arise if the features are spoiled and lose their representation during the calculations. Second, overfitting is the phenomenon of a large gap in model performance between training and inference data. As the model grows, the parameters may have more space to save in detail the features from the training dataset. However, there may be bias in some of the model parameters, which could have an impact on the results from which inferences are drawn. Several methods have been proposed to solve this problem, such as increasing the data volume, introducing more representative data, and appropriate feature selection.

A limitation of this study was insufficient data. Although the lower limit of the data volume acquired depends on tasks,
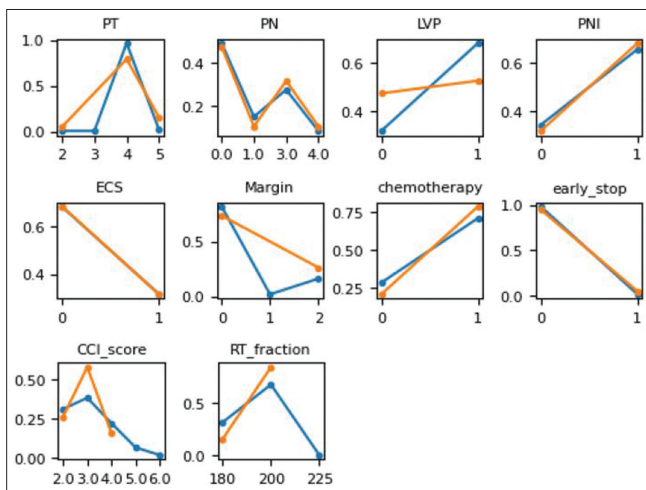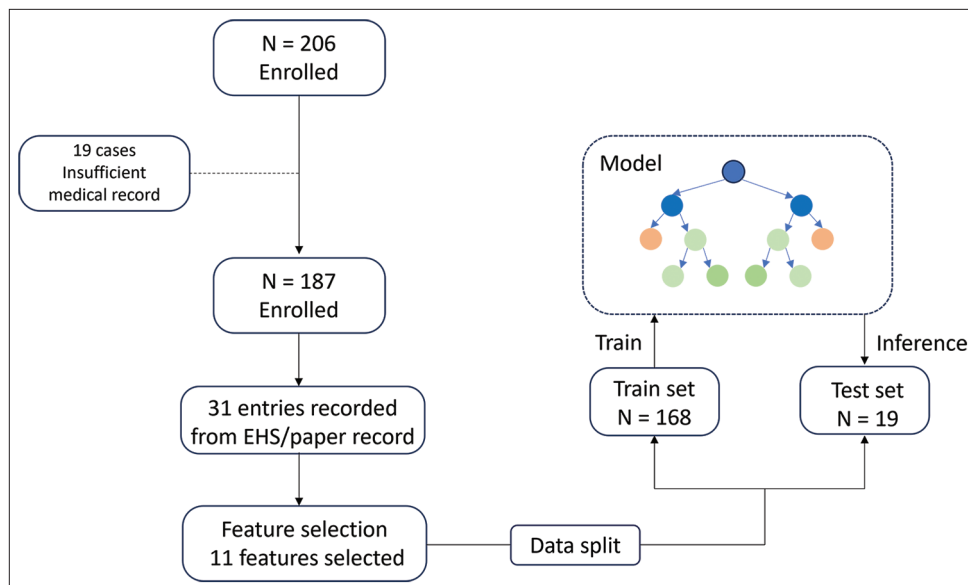

Figure 3: The data collection process. EHS: Electronic healthcare system

Table 4: The analysis of risk factors by regression model

Variables	Crude HR (95% CI)	*P*	Adjusted HR (95% CI)	*P*	Reference
Unfavorable group	1.95 (1.16−3.27)	0.011*	-	-	Favorable group
CCI score					
1	0.62 (0.37−1.03)	0.067			CCI score 0
2	1.15 (0.70−1.88)	0.578			
3	1.27 (0.70−2.30)	0.437			
4	0.30 (0.04−2.16)	0.230			
Differentiation	1.28 (0.79−2.07)	0.322	1.27 (0.77−2.09)	0.348	Differentiation (−)
LVI	1.47 (0.86−2.50)	0.158	1.35 (0.75−2.43)	0.314	LVI (−)
PNI	1.36 (0.81−2.89)	0.240	1.14 (0.64−2.01)	0.661	PNI (−)
Margin	0.96 (0.51−1.79)	0.897	0.87 (0.45−1.71)	0.694	Margin (−)
pN	2.30 (1.39−3.80)	0.001*	2.22 (1.21−4.05)	0.009*	pN (0)
ECS	1.74 (1.06−2.86)	0.029*	1.04 (0.55−1.98)	0.894	ECS (−)
CT	0.95 (0.56−1.72)	0.853	0.74 (0.40−1.36)	0.327	CT (−)

* *P*< 0.05, PNI: Perineural invasion, ENE: Extranodal extension, CCI: Charlson Comorbidity Index, ECS: Extracapsular spreading, CT: Computed tomography, HR: Hazard ratio, CI: Confidence interval, LVI: Lymphovascular invasion
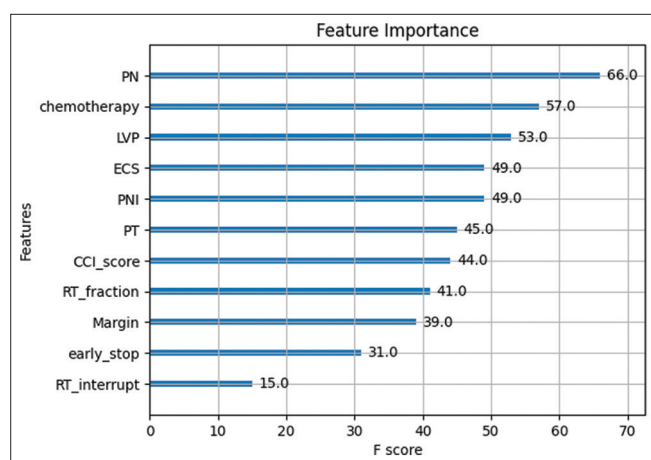

Figure 4: Feature importance in the model. PNI: Perineural invasion, ECS: Extracapsular spreading, LVP: Lymphovascular permeation, CCI: Charlson Comorbidity Index, PT: Pathological T stage, PN: Pathological N stage, RT: Radiotherapy

Figure 5: Receiver operating characteristic curve for the model of the testing dataset. ROC: Receiver operating characteristics

it is still difficult to construct a model for <200 cases. We tried to prune and aggregate features, so that the number of features would not overwhelm the model and cause overfitting. Nevertheless, the model yielded unsatisfactory results for the testing data. There may be two reasons for this problem: insufficient training data, as discussed above; and the low volume of testing data since they might not be representative enough, although we attempted to prove the matching of features between training and testing data [Figure 2].

The risk factors of patients with advanced head-and-neck cancer have been studied in postoperative settings. The RTOG 9501 and EORTC 22931 trials showed superior local control and overall survival in the chemoradiation group, especially in the case of ENE- and margin-involved tumors [3]. A previous study of a similar group of patients from Taiwan examined risk factors other than the two factors mentioned above. The authors aggregated several minor risk factors: pT4, pN1, close margin (<4 mm), poor differentiation of histology, PNI, lymphovascular permeation, or a tumor invasion depth of >1.1 cm. They found that patients with fewer than three minor factors had better outcomes, including better recurrence-free survival and overall survival [4].
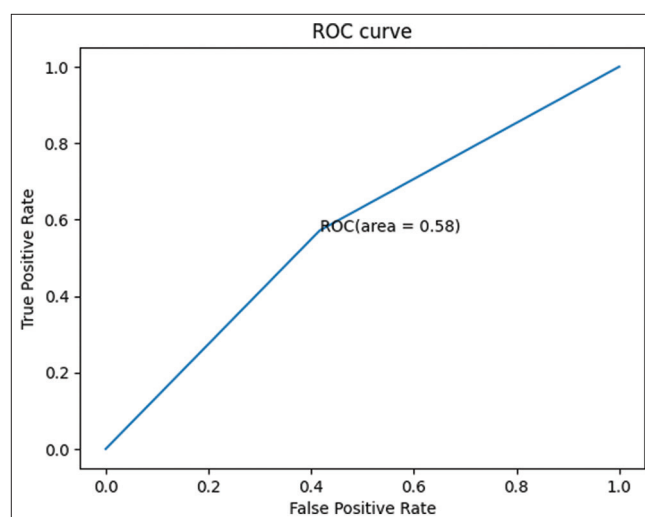
We attempted to determine the risk factors after the patients received adjuvant treatment. Figure 4 shows the feature importance of the model. Pathological lymph node status was the feature with the highest weighting. Pathological tumor status and risk factors were also important. These factors were primary pathology features. According to previous studies [3], they affect the differences in outcome between treatment groups.

We also took the factors associated with radiotherapy into consideration. Fraction size was a weighted feature in the model. Cumulative dose and fraction size are key treatment outcome factors in head-and-neck cancers [12-14]. At present, there is consensus on radiotherapy being the treatment for head-and-neck cancers, but physicians may adjust fraction size or dose to suit the patient's situation, which may lead to changes in treatment results and side effects. We established the treatment standard of radiotherapy between 2006 and 2007, following current recommendations from the guideline. Some patients enrolled were treated before that time; therefore, we assumed that these patients might have different outcomes.

Previous studies showed that interruption or even early termination of radiotherapy was critical to outcomes [15]. In our model, both interruption and early termination were weighted features. They might be correlated with other factors. For example, the more extensive the tumor-involved or surgery-involved region and the radiation field are, the more likely that radiation would cause adverse effects [16]. Age, comorbidity, and previous cancers, especially previous head-and-neck cancers, would also affect the treatment plan and patients' compliance. We integrated these factors into the CCI index [7], which was also a weighted feature in the model. However, there were other factors that might have led to treatment delay or interruption, such as tolerance to chemotherapy, patient or family support, or their decision to receive further treatment. Thus, we believe that they should be considered as separate independent factors for analysis.

Seventy-two percent of the patients enrolled received CCRT after primary surgery. Previous studies [3] have highlighted the importance of chemotherapy. It was a weighted feature in our model. Although patients with advanced-stage head-and-neck cancer tend to receive CCRT routinely at Hualien Tzu Chi General Hospital, we could follow the treatment outcomes of those not receiving CCRT due to future medical conditions.

The regression model identified similar potential risk factors to XGBoost. Both models defined pathological lymph node status as the most crucial factor that affected the treatment outcomes of patients. Pathological features such as lymphovascular permeation, ENE, PNI, surgical margin, and receiving chemotherapy or not were also critical to patients. It might suggest that whether a patient could have a longer time free from recurrence, or even a prolonged lifespan, was mainly decided by surgical performance and tumor behavior. However, more data are necessary to validate the trend.

This study has some limitations. The volume of data for model training was relatively small and there might have been bias in inferences drawn from the testing data. The cancer registry database was incomplete, especially for data before the 2010s, and data were also difficult to find in paper medical records. The staging system used for these patients was AJCC 7th edition, which is different from current practice because AJCC 8th edition has been applied since 2018. In the 8th edition, ENE is taken into account and becomes a critical feature for staging lymph node status [17]. This may affect the importance of features in the model.

Conclusions

We conducted an exploratory study for defining the risk factors of patients with advanced-stage head-and-neck cancer after adjuvant treatment. We used machine learning to develop a predictive model, and the potential risk factors found by the model had been approved by a regression model. We found that pathological staging, pathological risk factors, chemotherapy, and the quality of radiotherapy may affect the treatment outcome. However, the study had some limitations, including insufficient data volume, incomplete data registration, and differences in staging. In future studies, we hope to collect the data according to the features introduced in this study and to construct a stronger model to explain and predict the factors associated with treatment outcomes of patients with advanced-stage head-and-neck cancer.

References

1. Health Promotion Administration, Ministry of Health and Welfare. 2020 Cancer Registry Annul Report Taiwan; 2022.

2. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology, Head and Neck Cancers, Version 1; 2024.

3. Bernier J, Cooper JS, Pajak TF, van Glabbeke M, Bourhis J, Forastiere A, et al. Defining risk levels in locally advanced head and neck cancers: A comparative analysis of concurrent postoperative radiation plus chemotherapy trials of the EORTC (#22931) and RTOG (#9501). Head Neck 2005;27:843-50.

4. Fan KH, Chen YC, Lin CY, Kang CJ, Lee LY, Huang SF, et al. Postoperative radiotherapy with or without concurrent chemotherapy for oral squamous cell carcinoma in patients with three or more minor risk factors: A propensity score matching analysis. Radiat Oncol 2017;12:184.

5. Chen WC, Lai CH, Fang CC, Yang YH, Chen PC, Lee CP, et al. Identification of high-risk subgroups of patients with oral cavity cancer in need of postoperative adjuvant radiotherapy or chemo-radiotherapy. Medicine (Baltimore) 2016;95:e3770.

6. Sidey Gibbons JA, Sidey Gibbons CJ. Machine learning in medicine: A practical introduction. BMC Med Res Methodol 2019;19:64.

7. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. J Chronic Dis 1987;40:373-83.

8. Chang W, Liu Y, Xiao Y, Yuan X, Xu X, Zhang S, et al. A machine-learning-based prediction method for hypertension outcomes based on medical data. Diagnostics (Basel) 2019;9:178.

9. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J MLR 2011;12:2825-30.

10. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery; 2016. p. 785-94. doi: 10.1145/2939672.2939785.

11. Montomoli J, Romeo L, Moccia S, Bernardini M, Migliorelli L, Berardini D, et al. Machine learning using the extreme gradient boosting (XGBoost) algorithm predicts 5-day delta of SOFA score at ICU admission in COVID-19 patients. J Intensive Med 2021;1:110-6.

12. Horiot JC, Bontemps P, van den Bogaert W, Le Fur R, van den Weijngaert D, Bolla M, et al. Accelerated fractionation (AF) compared to conventional fractionation (CF) improves loco-regional control in the radiotherapy of advanced head and neck cancers: Results of the EORTC 22851 randomized trial. Radiother Oncol 1997;44:111-21.

13. Saksø M, Andersen E, Bentzen J, Andersen M, Johansen J, Primdahl H,

et al. A prospective, multicenter DAHANCA study of hyperfractionated, accelerated radiotherapy for head and neck squamous cell carcinoma. Acta Oncol 2019;58:1495-501.

14. Taylor JM, Mendenhall WM, Lavey RS. Dose, time, and fraction size issues for late effects in head and neck cancers. Int J Radiat Oncol Biol Phys 1992;22:3-11.

15. Thomas K, Martin T, Gao A, Ahn C, Wilhelm H, Schwartz DL. Interruptions of head and neck radiotherapy across insured and indigent patient populations. J Oncol Pract 2017;13:e319-28.

16. Grégoire V, Levendag P, Ang KK, Bernier J, Braaksma M, Budach V, et al. CT-based delineation of lymph node levels and related CTVs in the node-negative neck: DAHANCA, EORTC, GORTEC, NCIC, RTOG consensus guidelines. Radiother Oncol 2003;69:227-36.

17. Zanoni DK, Patel SG, Shah JP. Changes in the 8th edition of the American Joint Committee on Cancer (AJCC) Staging of head and neck cancer: Rationale and implications. Curr Oncol Rep 2019;21:52.