Received: 26 July 2022 | Revised: 27 April 2023 | Accepted: 10 June 2023

DOI: 10.1002/hed.27434

ORIGINAL ARTICLE

WILEY

Machine learning in laryngeal cancer: A pilot study to predict oncological outcomes and the role of adverse features

Gerardo Petruzzi MD[1] | **Elisa Coden MD**[2] | **Oreste Iocca MD, DDS**[3] | **Pasquale di Maio MD**[4] | **Barbara Pichi MD**[1] | **Flaminia Campo MD, PhD**[1] | **Armando De Virgilio MD, PhD**[5,6] | **Mazzola Francesco MD**[1] | **Antonello Vidiri MD**[7] | **Raul Pellini MD**[1]

[1]Department of Otolaryngology and Head and Neck Surgery, IRCCS Regina Elena National Cancer Institute, Rome, Italy

[2]Division of Otorhinolaryngology – Head and Neck Surgery, ASST Sette Laghi, Ospedale di Circolo e Fondazione Macchi, University of Insubria, Varese, Italy

[3]Division of Maxillofacial Surgery, Città della Salute e della Scienza, University of Torino, Torino, Italy

[4]Department of otolaryngology-Head and Neck Surgery, Giuseppe Fornaroli Hospital, ASST Ovest Milanese, Magenta, Italy

[5]Department of Biomedical Sciences, Humanitas University, Milan, Italy

[6]Department of Otolaryngology and Head and Neck Surgery, IRCCS Humanitas Research Hospital, Milan, Italy

[7]Department of Radiology and Diagnostic Imaging, IRCCS Regina Elena National Cancer Institute, Rome, Italy

Correspondence
Gerardo Petruzzi, Department of Otolaryngology-Head and Neck Surgery, IRCCS Regina Elena National Cancer Institute, Via Elio Chianesi 53, 00144, Rome, Italy.
Email: petruzzigerardo@gmail.com

Abstract

Background: Laryngeal carcinoma (LC) remains a significant economic and emotional problem to the healthcare system and severe social morbidity. New tools as Machine Learning could allow clinicians to develop accurate and reproducible treatments.

Methods: This study aims to evaluate the performance of a ML-algorithm in predicting 1- and 3-year overall survival (OS) in a cohort of patients surgical treated for LC. Moreover, the impact of different adverse features on prognosis will be investigated. Data was collected on oncological FU of 132 patients. A retrospective review was performed to create a dataset of 23 variables for each patient.

Results: The decision-tree algorithm is highly effective in predicting the prognosis, with a 95% accuracy in predicting the 1-year survival and 82.5% in 3-year survival; The measured AUC area is 0.886 at 1-year Test and 0.871 at 3-years Test. The measured AUC area is 0.917 at 1-year Training set and 0.964 at 3-years Training set. Factors that affected 1yOS are: LNR, type of surgery, and subsite. The most significant variables at 3yOS are: number of metastasis, perineural invasion and Grading.

Conclusions: The integration of ML in medical practices could revolutionize our approach on cancer pathology.

KEYWORDS

algorithm, artificial intelligence, laryngeal cancer, machine learning, oncological outcome, open surgery

1 | INTRODUCTION

Laryngeal cancer (LC) is one of the most common malignant tumors of the head and neck region. It accounts for 177 000 new cases per year worldwide, with 94 000 related deaths.[1,2] Major risk factors include tobacco, smoking, and alcohol consumption. Thanks to anti-smoking campaigns, tobacco use has declined over the past two decades, resulting in a 2.4% per year decrease in the incidence of LC. On the other hand, the prognosis has not improved significantly, with a 5-years overall survival (OS) of 60.9%.[3]

Tumor stage at diagnosis impacts on OS and it largely varies from 85% at early-stage to 40% in stage IV disease.[4] Due to this discrepancy, many efforts have been made to stratify patients into high- or low-risk groups and to plan patient-specific optimal treatments.[5–8]

The collaboration between biomedicals and bioinformatics has led to develop the Artificial Intelligence (AI) algorithms, which have consistently improved the cancer's analysis compared with conventional statistical tools.[9–11] Machine learning (ML) is a branch of AI that deals with predictions in unknown situations from previous observations that have been "learnt."[12–15] The application of ML in medical practice can potentially revolutionize patient care and therapeutic decisions by individualizing the treatment according to specific risk-factors.[16,17]

Recently the applications of AI in LC have encompassed a variety of fields, including radiomics, genomics, acoustics, and videomics to support screening, diagnosis, decision making, and oncological outcome.

However, the common drawback of these studies lies in their verification, since most of them remain in a proof-of-concept level, because they are bound to databases of singular institutions or address only a few limited topics. To date, the use of AI in clinical practice is absent.

The most developed areas are: videomics (AI can classify LC, distinguish precancerous lesions, benign lesions, or healthy patients learning from laryngoscopy images); acoustic data evaluation (the voice signal analysis can be used to distinguish patients with laryngeal cancer from healthy subjects). In other fields (e.g., radiomics) AI may be capable to distinguish benign from malignant lesions or automatically carry out diagnosis analyzing histopathologic slides.[13,14]

As regards management and the outcomes (risk of recurrence, possibility of distant metastases, therapeutic choice) the use of artificial intelligence is still in its initial stages.[17]

This pilot study aims to evaluate the performance of a ML algorithm in predicting 1- and 3-year OS in a cohort of patient surgical treated for LC. Moreover, the role and the impact of different adverse features will be investigated with a decisional tree.

Data are collected on surgical experiences and oncological follow-up of 132 patients affected by laryngeal cancer, treated at IRCCS "Regina Elena" National Cancer Institute of Rome.

We kindly encourage colleagues to co-operate with our project by sharing database or testing patients with our algorithm. The algorithm is available on: www.datalarynx.com.

2 | MATERIALS AND METHODS

2.1 | Endpoint of the study

The primary endpoint of the study is the prediction of the oncological outcome at two different time points (1- and 3-years after surgery). The oncological outcomes are defined as: (a) NED (No evidence of disease); (b) RD (Relapse of disease). The latter includes both AWD (Alive with Disease) and DOD (Dead of Disease) patients.

A secondary endpoint is the impact of the different adverse features in predicting the oncological outcome.

2.2 | Patient selection and dataset creation

All patients who underwent open surgical treatment for laryngeal squamous cell carcinoma (SCC) at the IRCCS "Regina Elena" National Cancer Institute between January 1, 2005 and August 31, 2021 were enrolled. The Institutional Review Board and the Ethics Committee of our institution approved the study. (Protocol No. RS1661/22 IFO).

Inclusion criteria were: (1) histologically confirmed SCC of the larynx[18]; (2) indication of the Disease Management Team to an upfront open surgical treatment according to Head and Neck Cancers NCCN clinical practice Guidelines[19]; (3) complete follow-up of at least 3 years.

Exclusion criteria were: (1) patients who had undergone previous surgical or radio-chemotherapy treatment on the head and neck region; (2) patients with synchronous or metastatic tumors; (3) patients without concomitant neck dissection; (4) incomplete data records or "lost to follow-up"; and (5) patients dead for other causes.

Enrolled patients were defined as "instances." A retrospective review of clinical and pathological records was performed to create a dataset containing a total of 23 variables, called "attributes," which are listed in Table 1. Exactly 1- and 3-year OS were established as the "target" attributes to be predicted.

TABLE 1 List of attributes.

	Attribute name	Attribute values	definition
1	Age at surgery	yy (*years*)	Patient age at the time of surgery
2	Gender	male (M); female (F)	Patient biological sex assigned at birth
3	Smoke	yes; no	Previous or active tobacco smoking habit for more than 1 year
4	Pack-years	nn (pack-years)	Unit of measure for tobacco exposure (20-cigarette packs smoked per day multiplied by the number of years the person has smoked)
5	Alcohol	Yes; no	For men, consuming more than 4 drinks on any day or more than 14 drinks per week. For women, consuming more than 3 drinks on any day or more than 7 drinks per week
6	Cancer subsite	Supraglottis; glottis; subglottis	Anatomical subsite of cancer origin within the larynx
7	Treatment	Surgery alone; Surgery + adjuvant RT; Surgery + concomitant RT/CHT	Treatment modality
8	Type of laryngectomy	Total laringectomy; OPHL 1; OPHL 2A; OPHL 2B; OPHL 3	Type of laryngectomy performed in patients undergoing surgical treatment (according to the European Laryngological Society).
9	Type of neck dissection	SND, mRND, RND; SND + SND; SND + mRND; SND + RND; mRND + mRND; mRND + RND; RND + RND	Type of neck dissection performed in patients undergoing surgical treatment (according to the American Academy of Otolaryngology). When bilateral, the side homolateral to the cancer is indicated first
10	Surgical Margins	R1, R2, Close (<2 mm), R0	The distance between the invasive tumor front to the resected margin
11	Grading	1; 2; 3	Histologic grading according to Broder classification[5]
12	pT	1A; 1B; 2; 3; 4A; 4B	Pathologic assessment of the primary tumor (T category) according to AJCC Cancer Staging Manual, 8th Ed
13	vI	0; 1	Histologic evidence of vascular infiltration
14	nI	0; 1	Histologic evidence of perineural infiltration
15	pN	0; 1; 2A; 2B; 2C; 3A; 3B	Pathologic assessment of lymph node involvement (N category) according to AJCC Cancer Staging Manual, 8th Ed
16	ENE	0; 1	Histologic evidence of extranodal extension
17	cM	0; 1	Clinical assessment of lymph node involvement (N category) according to AJCC Cancer Staging Manual, 8th Ed
18	Staging	I; II; III; IVA; IVB; IVC	Prognostic stage group according to AJCC Cancer Staging Manual, 8th Ed
19	Total number of lymph nodes removed	nn	Total number of lymph nodes isolated from the surgical specimen
20	Number of metastatic lymph nodes	NN	Number of metastatic lymph nodes isolated from the surgical specimen
21	LNR	NN/nn	Lymph node ratio (ratio between number of metastatic lymph nodes and total number of lymph nodes isolated from the surgical specimen)
22	1-year follow-up	NED; RD	Oncological status at 1-year follow-up NED = No evidence of disease); RD Relapse of disease (AWD = alive with disease or DOD dead of disease)
23	3-year follow-up	NED; RD	Oncological status at 1-year follow-up NED = No evidence of disease; RD Relapse of disease (AWD = alive with disease or DOD = Dead of disease)

2.3 | Algorithm

Waikato Environment for Knowledge Analysis (Weka) was used as a data mining tool.[20] Weka is an open source software developed at the University of Waikato, New Zealand, and licensed under the GNU General Public License. Weka provides users with a set of algorithms for data mining tasks, including preprocessing, classification, regression, clustering, association rules, feature selection, and visualization.[21–23] For this study we chose the J48 algorithm in the Weka Data mining tool, an open-source java implementation of the C4.5 algorithm, whose output is used to generate a decision tree.[24,25]

2.4 | Data processing

A series of preprocessing operations were carried out on the raw data to optimize the information extraction process by reducing noise, redundancy and waste. Missing values, outliers (e.g., abnormal values), and useless data were removed from the dataset ("data cleaning"); continuous values were normalized and all the attributes were generalized according to concept definitions ("data transformation"). Only a subset of highly informative attributes was selected for the next training phase. This step was performed using the Correlation-based Feature Selection (CFS), a WEKA function.

2.5 | Classification

The dataset was randomly organized to ensure a casual insertion of the instances (the patients) and to avoid bias. This randomization was carried out through an additional algorithm that generated a casual distribution. The additional algorithm started from an initial numeric value which is called "seed."

In the ML algorithm, the dataset was split into two parts: training data used to build the model and test data used to evaluate the performance of the algorithm and to characterize the generalization capabilities of the model. We opted for a percentage split: 70% of the instances (e.g., 92/132 patients) were used as training data and the remaining 30% (e.g., 40/132 patients) as test data.

This operation was repeated iteratively on the dataset with multiple input seeds: we iteratively subjected to the algorithm a different distribution of the data.

2.6 | Metrics and evaluation

To evaluate the goodness and the reliability of the model obtained by the algorithm, we used metrics and indicators for binary classifiers.

The basic metrics were: True Positive (TP): number of correctly classified positive examples; False Negative (FN): number of positive examples erroneously classified as negative; False Positive (FP): number of negative examples erroneously classified as positive; True Negative (TN): number of correctly classified negative examples.

A confusion matrix was created to represent false negatives and false positives produced by the model. Starting from these values, more complex indicators were extracted: Accuracy; Recall or TPR (True Positive Rate); Specificity or TNR (True Negative Rate); False Positive Rate; False Negative Rate; Precision; F-measure; MCC (Matthews correlation coefficient); Kappa statistics (Cohen's Kappa) with the Landis and Koch ranges (Kappa <0: there is no agreement; Kappa $\in$ (0; 0.2): weak; Kappa $\in$ (0.2; 0.4): enough; Kappa $\in$ (0.4; 0.6): good; Kappa $\in$ (0.6; 0.8): excellent; Kappa $\in$ (0.8; 1): almost perfect).

3 | RESULTS

3.1 | Clinical dataset

A total of 132 on 673 patients who underwent open surgical treatment for laryngeal SCC at our institution met the inclusion criteria.

A total of 107 of them were males (81.06%), 25 were females (18.94%) with a male to female ratio of 5.28:1. The mean age at treatment was 62 years (range 19–88 years). Exactly 98 were active or former smokers (range 0–140 pack-years) while 44 reported alcohol abuse. All the alcohol abusers were also smokers. The most affected laryngeal subsite was the glottis (94 cases), followed by the supraglottis in 37 cases and the subglottis in 1 case.

All patients underwent upfront surgical treatment: 8 patients were eligible for Open partial horizontal laryngectomy (OPHL) type I; 31 patients were treated with OPHL type IIa; 15 patients with OPHL type IIb, 5 patients for OPHL type III; 73 patients were treated with total laryngectomy.[26] All patients received neck dissection (ND): 43 patients received monolateral selective neck dissection (SND); 61 patients received bilateral SND; 10 patients were treated with bilateral radical neck dissection; 18 patients received monolateral modified radical neck dissection and contralateral SND.

In all cases, the histological examination confirmed the diagnosis of SCC. According to Broder classification, 2 out of the 132 patients were G1, 67 G2, and 63 G3.

All patients were re-staged. According to AJCC TNM 8th Ed,[27] six patients were staged as pT1, 23 as pT2, 55 as pT3, and 48 as pT4. Neck dissection specimens were pathologically negative (pN0) in 88 patients, pN1 in 12 cases, pN2b in 12 cases, pN2c in 3 cases, while 17 patients were classified as pN3b due to extracapsular spread (ENE+).

The number of lymph nodes removed per patient ranged from 12 (monolateral ND) to 121 (bilateral ND). The number of lymph node metastases per patient varied from 0 (N0) to 28 (N+). The mean Lymph Node Ratio (LNR—the number of positive lymph nodes divided by the total number of lymph nodes excised) was 0.014 (range 0–0.298).[28]

As regarded post-operative Stage: 6 patients were Stage I, 24 Stage II, 41 Stage III, 43 Stage IVA, and 18 Stage IVB. Concerning post-operative adverse features, vascular infiltration was found in 18 cases; perineural invasion in 23 cases. Adjuvant radiotherapy was administered in 56 patients. Eleven patients received adjuvant radio-chemotherapy. All patients were re-evaluated at 1- and 3-years.

At 1-year follow-up, 114 patients were alive without evidence of disease (NED). Eight adverse events had occurred: 12 patients had recurrence or persistence (AWD); 6 deaths due to disease-related causes (DOD).

At 3-year follow-up, 103 patients were NED; 11 patients were AWD; 18 patients were DOD. Compared with the previous endpoint, 11 additional adverse events were observed (9 relapses and 2 disease-related deaths in patients who were NED at 1 year).

The main demographic, surgical and clinical characteristics were summarized in Table 2.

3.2 | Machine learning evaluation

The dataset has been randomly ordered with an additional algorithm, starting from a seed.

It was preferred the seed-generated random distribution that provided a homogeneous distribution of the data in terms of training and testing.

This operation preceded the evaluation of the model obtained with the machine learning algorithm. Non-homogeneous distributions were also tested, confirming the inadequacy of the results: even if they presented very high accuracy values, their Kappa statistic were close to zero.

3.3 | Evaluation at 1-year

Table 3A presents the results obtained from seed 22. The training work was carried out on 92 patients of which 79 were NED and 13 with adverse events (AWD or DOD), the testing one on the remaining 40 of which 35 were NED and 5 labeled as AWD or DOD. A total of 38 out of 40 patients were correctly classified, obtaining an accuracy of 95%.

As shown in Table 3A, a True Positive Rate close to 1 demonstrates the ability of the classifier to identify positive results, in the same way the Precision and Recall which close to 1 shows low false positives and false negatives, respectively. Similarly, F-measure and MCC return high scores. The Kappa statistic has a value of 0.7714, with an agreement level defined as excellent.

The measured AUC area is 0.886 in the Testing set and it counts 0.917 in the Training one.

Regarding the importance of attributes, it is necessary to analyze the decision tree produced with the training data (Figure 1).

The root of the tree considered the most important prognostic factor is the LNR. All patients with a LNR included in the range 0–0.03 are NED at 1-year.

The left branch with low values of LNR contains 71 patients, equal to 77.2% of the total training data (92 cases). 69/71 patients are correctly classified. Therefore 97% of 77.2% of the training data is correctly classified with LNR only.

Three other prognostic factors are elected on the right branch of the tree. The outcome of patients with a high LNR value is influenced by type of surgery on T, type of ND and laryngeal subsite. The supraglottic laryngeal cancer or a surgical treatment with OPHL type 3 seems to have a high risk of RD (3 out of 4 patients with those variables are recognized by the algorithm and correctly classified as patients with relapse of the disease). All other variables have apparently no impact on prognosis at 1-year follow-up.

3.4 | Evaluation at 3-year

The same approach was used to predict the mentioned endpoint at 3-years after surgery. In Table 3B we present the results obtained for seed 51, a seed with a good distribution of data between training and testing.

The algorithm ran training work on 92 patients (70% of the cohort) of which 71 were NED and 21 with adverse events (AWD or DOD), then it tested itself with the remaining 40 of which 32 were NED and 8 labeled as AWD or DOD. Exactly 33 out of 40 patients were correctly classified, achieving an accuracy of 82.5%. A true positive rate close to 1 demonstrates the ability to identify positive results, similarly the Precision and Recall close to 1 shows low false positives and false negatives respectively. The Kappa statistic has a value of 0.557, with a level of agreement defined as good. The measured AUC area is 0.871 in the Testing set and it counts 0.964 in the Training one.

TABLE 2 Demographic, surgical, and oncological information.

Variable	Distribution	Data
Gender	Male	107 (81.1%)
	Female	25 (18.9%)
Age (years)	Mean (Range)	62 (19–88)
Subsite	Glottis	94 (71.2%)
	Supraglottis	37 (28.0%)
	Subglottis	1 (0.7%)
Type of laryngectoymy	OPHL type I	8 (6.1%)
	OPHL type IIa	31 (23.5%)
	OPHL type IIb	15 (11.4%)
	OPHL type III	5 (3.8%)
	Total laryngectomy	73 (55.3%)
Type of neck dissection	Homolateral SND	43 (32.6%)
	Bilateral SND	61 (46.2%)
	Homolateral mRND + contralateral SND	18 (13.7%)
	Bilateral mRND	10 (7.6%)
pT stage	pT1	6 (4.5%)
	pT2	23 (17.4%)
	pT3	55 (41.7%)
	pT4	48 (36.4%)
pN stage	pN0	88 (66.7%)
	pN1	12 (9.1%)
	pN2b	12 (9.1%)
	pN2c	3 (2.3%)
	pN3b	17 (12.9%)
Stage	I	6 (45.4%)
	II	24 (18.2%)
	III	41 (31.1%)
	IVa	43 (32.6%)
	IVb	18 (13.7%)
Grading	G1	2 (1.5%)
	G2	67 (50.8%)
	G3	63 (47.7%)
N	ENE+	17 (12.9%)
Surgical margins	R0	132 (100%)
nI	Perineural invasion	23 (17.4%)
vI	Vascular invasion	18 (13.6%)
LNR	Mean (range)	1.4% (0–29.8%)
Adjuvant treatment	Radiation therapy	56 (42.4%)
	Chemo-radiation therapy	11 (8.3%)
	None	65 (49.2%)
1-year follow up	NED	114 (86.4%)
	AWD	12 (9.1%)
	DOD	6 (4.5%)
3-year follow up	NED	103 (78.0%)
	AWD	11 (8.3%)
	DOD	18 (13.7%)

TABLE 3 Evaluation results—(A) evaluation at 1-year; (B) evaluation at 3-year.

A)	TP rate	FP rate	Precision	Recall	*F*-measure	MCC	AUC area	Class
Weighted avg	0.800	0.029	0.800	0.800	0.800	0.771	0.886	RD
	0.971	0.200	0.971	0.971	0.971	0.771	0.886	NED
	0.950	0.179	0.950	0.950	0.950	0.771	0.886	
					Unit		Percentage	
	Correctly classified instances				38		95%	
	Incorrectly classified instances				2		5%	
	Kappa statistic				0.7714		77.14%	
B)	**TP rate**	**FP rate**	**Precision**	**Recall**	***F*-measure**	**MCC**	**AUC area**	**Class**
Weighted avg	0.875	0.188	0.538	0.875	0.667	0.587	0.871	RD
	0.813	0.125	0.963	0.813	0.881	0.587	0.871	NED
	0.825	0.138	0.878	0.825	0.838	0.587	0.871	
					Unit		Percentage	
	Correctly classified instances				33		82.5%	
	Incorrectly classified instances				7		17.5%	
	Kappa statistic				0.557		55.7%	

The performance of the classifier gets worse compared with the previous scenario, 17.5% of the data is classified incorrectly.

Figure 2 represents the decision tree obtained. The root of the tree is the presence of lymph node metastases. A total of 43/92 patients (45.65%) lied in the left branch of the tree. Their outcome was "guessed" by the algorithm at 97.67%. Their outcome depends on three different attributes: the number of lymph node metastases, perineural invasion, and Grading. All patients with 0 or 1 lymph node metastases, no perineural invasion, and Grade 1 or 2 were NED at 3 years of follow-up. In case of different values (e.g., perineural invasion, 2 or more nodes, G3) the risk of RD increases. The tree has three other leaves influencing outcome: type of surgery on T, type of ND, pN. However, the absence of consistent data did not ensure the accuracy of last three attributes.

4 | DISCUSSION

LC is a relatively rare neoplasm still associated with high mortality and morbidity.[1,2] In addition to early diagnosis, a better comprehension of the most important prognostic factors can contribute to develop a more personalized treatment plan.[7,16]

Currently, LC treatment is based on the presence of certain clinical or histological "adverse features," such as: advanced cTNM stage, status of the margin, or lymph node extracapsular extension.[5,19] Evidence is mainly based on empirical experiences and outdated scientific studies with heterogeneous or small cohorts.[29–31]

Developments in statistics and computer engineering over the last two decades have encouraged many scientists to apply statistical methods to analyze oncological outcomes or to offer important predictive information. To improve the evidence level, researchers are recently focusing on Artificial Intelligence.[32,33] The AI-applications in cancer diagnostics, prognosis predictions, and trials represent a current challenge in the oncologic area.[34,35]

As regards the optimal machine learning algorithm to be adopted, no consensus has been expressed in scientific community; the choice is generally based on: quantity and quality of data, aim of study, previous reports in the literature, and investigator's experience.[32]

The aim of this work is not to compare different types of algorithms, but to deep the decision role that leads to an optimized oncological prediction. The chosen algorithm is J48, an open-source Java implementation of the C4.5 algorithm in the Weka data mining tool.[20,24] Its output is a decision tree, a graphical tree-like structured classifier with root node (e.g., the person has lymph node metastasis or not) and terminal leaves determined by the answers. Algorithms travel from a root node to a leaf node based on learning rules derived from data training.[25] For medical purposes a decision tree was considered the most suitable choice, as it provides readable rules easy to be understood and built.[23]

In our study, we collected 23 attributes related to the patients, the type of tumor and the type of treatment. We asked the machine to learn the patient's 1- and 3-year prognosis in terms of NED, AWD or DOD and then to predict the outcomes in a subset of patients. As a second goal we evaluated through a decisional tree which

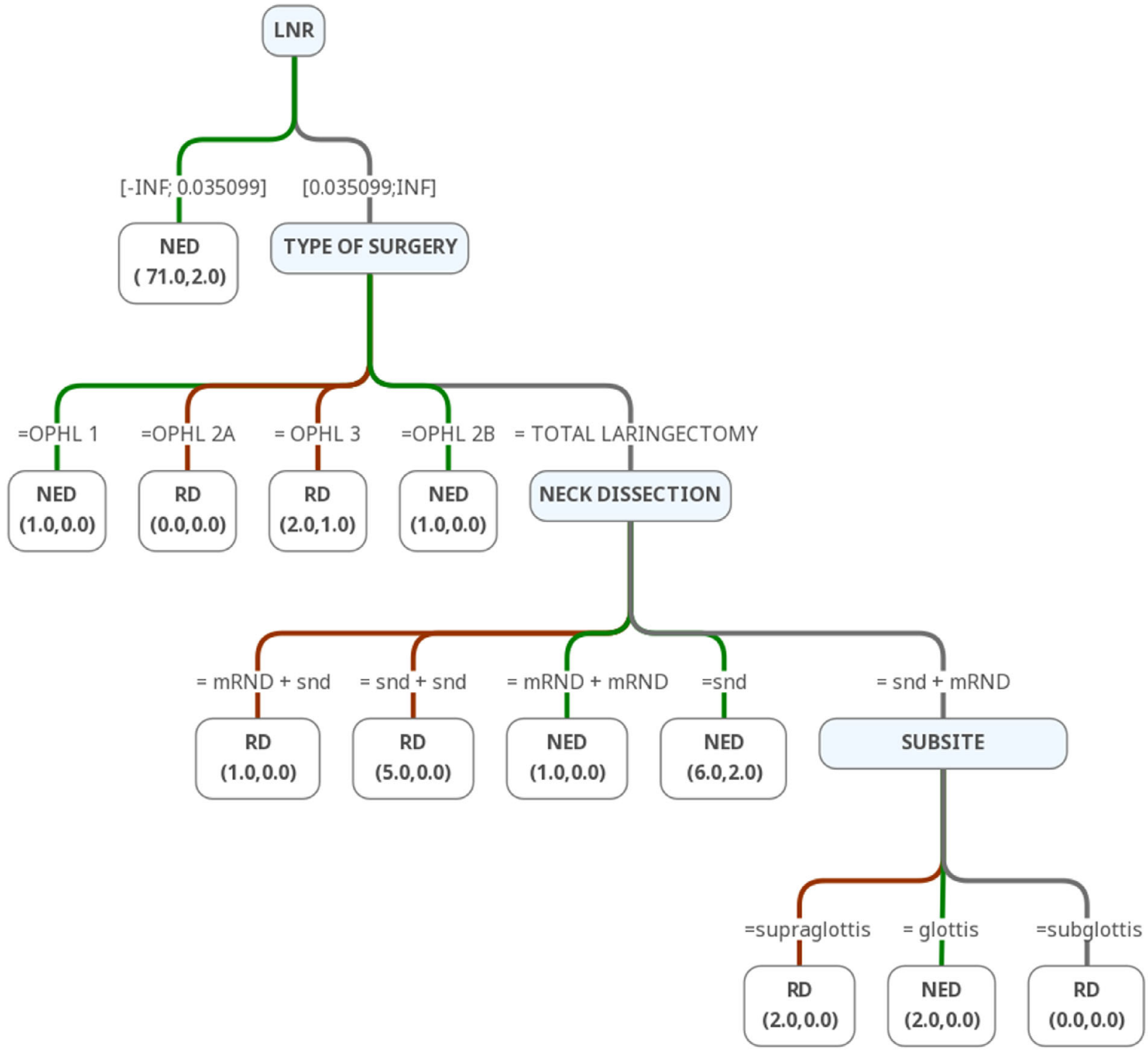

FIGURE 1 Decision tree at 1 year. The branches of the tree were colored according to the class (red in case of adverse events, green in case of a good outcome; the gray branches are classifications not supported by data). [Color figure can be viewed at wileyonlinelibrary.com]

attributes have a greater impact on prognosis. Our results demonstrate the high effectiveness of the algorithm in predicting the prognosis of laryngeal cancer, with a 95% accuracy in predicting the 1-year survival and 82.5% in 3-year survival. According to our algorithm, LNR, type of surgery on T, type of ND, and cancer subsite were the key factors that affected 1-year overall survival. The most significant variables in predicting 3-year overall survival were: number of lymph node metastasis, presence of perineural invasion, and grading. Therefore, an advanced tumor with a high number of metastases at the time of diagnosis has an important impact on the short-term prognosis at 1-year. On the other hand, the intrinsic characteristics of the neoplasm, such as the grading or the capacity of perineural infiltration, determine a low long-term prognosis influencing the 3-year outcome.

The idea of a better identification of prognostic factors and related influence could have important reverberations on clinical practice and our daily therapeutic choices.

Looking at the results obtained with our J48 algorithm, we re-evaluated our database with a machine learning from another family of algorithm. The comparison seemed necessary to see if the results changed using different algorithms. We have added to the evaluation with the J48 algorithm also the results obtained with the Naive Bayes of WEKA.[22] Naive Bayes classifiers are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong (naive) independence assumptions between the features.[36]

Here follow the results obtained at 1-year. The training work was carried out on 92 patients. Correctly

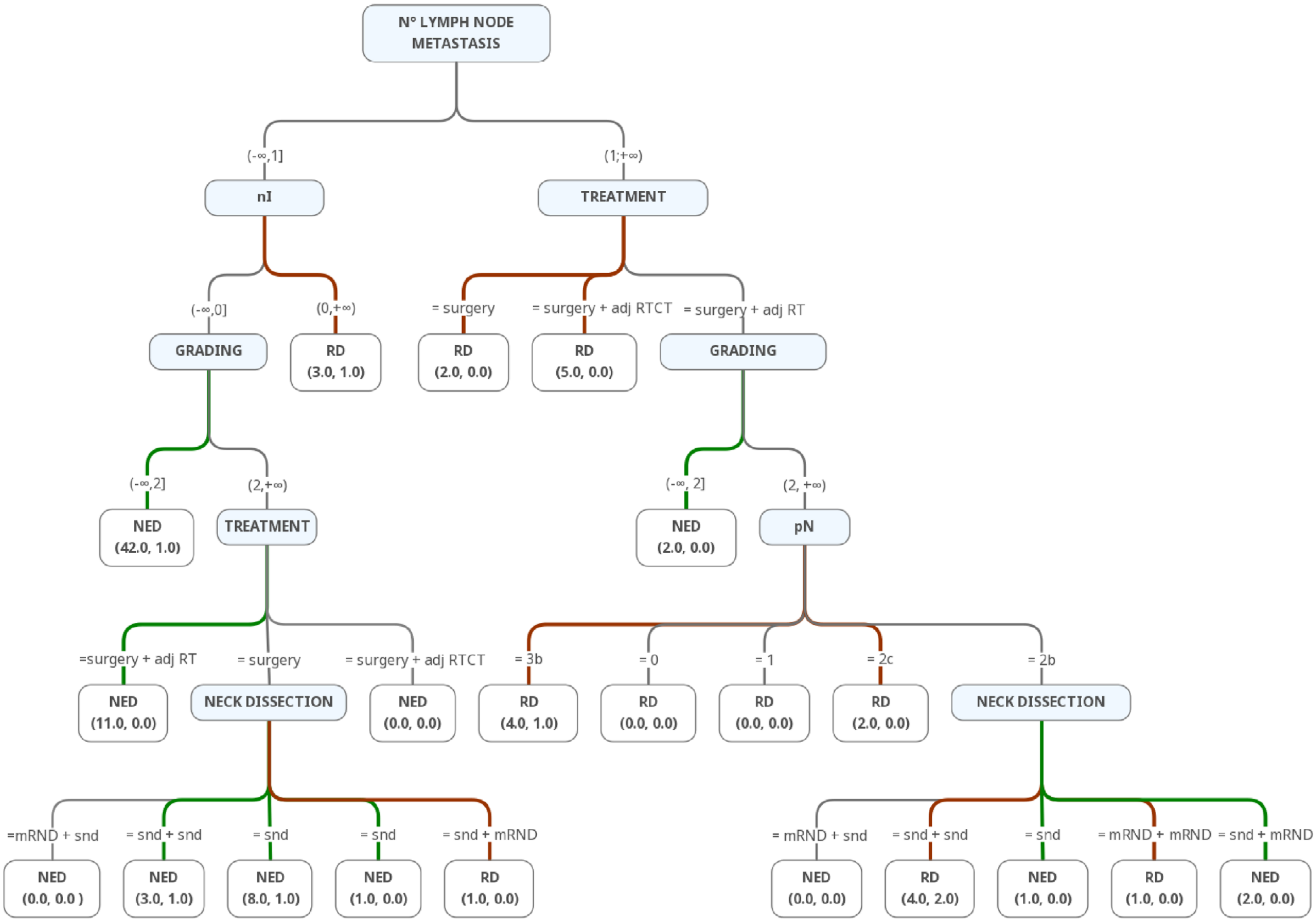

FIGURE 2 Decision tree at 3 years. The branches of the tree were colored according to the class (red in case of adverse events, green in case of a good outcome; the gray branches are classifications not supported by data). [Color figure can be viewed at wileyonlinelibrary.com]

classified instances were 78 (84.7826%); Incorrectly classified instances were 14 (15.2174%); AUC area was 0.923. The testing work was carried out on the remaining 40 patients. Correctly classified instances were 34 (85%); Incorrectly classified instances were 6 (15%); AUC area was 0.831. Results obtained at 3-years. The algorithm ran training work on 92 patients. Correctly classified instances were 78 (84.7826%); Incorrectly classified instances were 14 (15.2174%); AUC area was 0.888. Then the algorithm tested itself with the remaining 40. Correctly classified instances were 35 (87.5%); Incorrectly classified instances were 5 (12.5%); AUC area was 0.898. The results are statistically comparable to those obtained with the J48 algorithm. This demonstrates that other algorithm models can also be used, proving to be equally effective and capable of producing comparable results.

The present study has some limitations. First, the machine learning algorithms were developed using a retrospective cohort, limited data, and selected attributes. Second, this study is based on patients treated in a single institution by the same surgical équipe and evaluated by the same multidisciplinary team. In addition, AI may be affected by the so-called "black box effect": researchers and clinicians typically know the inputs and the results, but it is hard to understand what is going on inside of the algorithm. The relative absence of transparency continues to be a limitation and constitutes a lack of confidence in results. Finally, many ethical questions remain unanswered: what about the patient's reactions in the event of an unfavorable prediction? Are they less likely to continue the treatment in case of a low prognosis? Who is responsible for an incorrect prediction?

5 | CONCLUSION

Laryngeal cancer continues to be a significant economic alarm bell for the healthcare system and an equally important social problem for patients and caregivers.

Traditionally, the experience of radiotherapists, radiologists, and surgeons plays a significant role in therapies decision-making; this historical approach, despite with the diffusion of various guidelines, poses a great risk of bias and a subjective choice of the treatment.

Nowadays AI is a rapidly evolving field. The integration of AI into our daily routine and communication has defined the Fourth Industrial Age and may renew our approach to cancer pathology. In this pilot study, we demonstrate the application of machine learning using our institutional database. We developed a decision-tree algorithm which is highly effective in predicting the prognosis of laryngeal cancer, with a 95% accuracy in 1-year survival and 82.5% in 3-year survival.

We invite the scientific community to contribute to our database and to validate our data, making them more efficient. Follow our website to test the algorithm, predicting your patient's outcomes and contribute to a future and shared database (www.datalarynx.com).

ACKNOWLEDGMENTS

The authors are grateful to Dr. Antonella Botte, Software Engineer and Dr. Claudio Moretti, MD, ENT Resident.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author. Raw Data are submitted to the Journal.

ORCID

Gerardo Petruzzi https://orcid.org/0000-0003-0371-7667
Oreste Iocca https://orcid.org/0000-0002-4444-248X
Barbara Pichi https://orcid.org/0000-0003-0955-4563
Armando De Virgilio https://orcid.org/0000-0003-0738-8223
Raul Pellini https://orcid.org/0000-0001-6051-3041

REFERENCES

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209-249. doi:10.3322/caac.21660
2. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin*. 2022;72(1):7-33. doi:10.3322/caac.21708
3. Hoffman HT, Porter K, Karnell LH, et al. Laryngeal cancer in the United States: changes in demographics, patterns of care, and survival. *Laryngoscope*. 2006;116(9):1-13. doi:10.1097/01.mlg.0000236095.97947.26
4. Harris BN, Bhuskute AA, Rao S, Farwell DG, Bewley AF. Primary surgery for advanced-stage laryngeal cancer: a stage and subsite-specific survival analysis. *Head Neck*. 2016;38(9):1380-1386. doi:10.1002/hed.24443
5. Machiels JP, René Leemans C, Golusinski W, Grau C, Licitra L, Gregoire V. Squamous cell carcinoma of the oral cavity, larynx, oropharynx and hypopharynx: EHNS-ESMO-ESTRO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2020;31(11):1462-1475. doi:10.1016/j.annonc.2020.07.011
6. Muscatello L, Piazza C, Peretti G, et al. Open partial horizontal laryngectomy and adjuvant (chemo)radiotherapy for laryngeal squamous cell carcinoma: results from a multicenter Italian experience. *Eur Arch Otorhinolaryngol*. 2021;278(10):4059-4065. doi:10.1007/s00405-021-06651-6
7. Birkeland AC, Beesley L, Bellile E, et al. Predictors of survival after total laryngectomy for recurrent/persistent laryngeal squamous cell carcinoma. *Head Neck*. 2017;39(12):2512-2518. doi:10.1002/hed.24918
8. Petruzzi G, Costantino A, De Virgilio A, et al. Lymphoepithelial carcinoma of larynx and hypopharynx: a systematic review and pooled analysis. *Eur Arch Otorhinolaryngol*. 2022;279(3):1157-1166. doi:10.1007/s00405-021-07011-0
9. Karadaghy OA, Shew M, New J, Bur AM. Development and assessment of a machine learning model to help predict survival among patients with oral squamous cell carcinoma. *JAMA Otolaryngol Head Neck Surg*. 2019;145(12):1115-1120. doi:10.1001/jamaoto.2019.0981
10. Smith JB, Shew M, Karadaghy OA, et al. Predicting salvage laryngectomy in patients treated with primary nonsurgical therapy for laryngeal squamous cell carcinoma using machine learning. *Head Neck*. 2020;42(9):2330-2339. doi:10.1002/hed.26246
11. Cavalieri S, De Cecco L, Brakenhoff RH, et al. Development of a multiomics database for personalized prognostic forecasting in head and neck cancer: the big data to decide EU project. *Head Neck*. 2021;43(2):601-612. doi:10.1002/hed.26515
12. Crowson MG, Ranisau J, Eskander A, et al. A contemporary review of machine learning in otolaryngology-head and neck surgery. *Laryngoscope*. 2020;130(1):45-51. doi:10.1002/lary.27850
13. Xiong H, Lin P, Yu JG, et al. Computer-aided diagnosis of laryngeal cancer via deep learning based on laryngoscopic images. *EBioMedicine*. 2019;48:92-99. doi:10.1016/j.ebiom.2019.08.075
14. Kim H, Jeon J, Han YJ, et al. Convolutional neural network classifies pathological voice change in laryngeal cancer with high accuracy. *J Clin Med*. 2020;9(11):3415. doi:10.3390/jcm9113415
15. Howard FM, Kochanny S, Koshy M, Spiotto M, Pearson AT. Machine learning-guided adjuvant treatment of head and neck cancer. *JAMA Netw Open*. 2020;3(11):e2025881. doi:10.1001/jamanetworkopen.2020.25881
16. Kumar S, Patnaik S, Dixit A. Predictive models for stage and risk classification in head and neck squamous cell carcinoma (HNSCC). *PeerJ*. 2020;22(8):e9656. doi:10.7717/peerj.9656
17. Bensoussan Y, Vanstrum EB, Johns MM 3rd, Rameau A. Artificial intelligence and laryngeal cancer: from screening to prognosis: a state of the art review. *Otolaryngol Head Neck Surg*. 2023;168(3):319-329. doi:10.1177/01945998221110839
18. Gale N, Poljak M, Zidar N. Update from the 4th edition of the World Health Organization classification of head and neck tumours: what is new in the 2017 WHO blue book for tumours of the hypopharynx, larynx, trachea and parapharyngeal space. *Head Neck Pathol*. 2017;11(1):23-32. doi:10.1007/s12105-017-0788-z

19. Caudell JJ, Gillison ML, Maghami E, et al. NCCN guidelines® insights: head and neck cancers, version 1.2022. *J Natl Compr Canc Netw*. 2022;20(3):224-234. doi:10.6004/jnccn.2022.0016
20. Kumar N, Khatri S. Implementing WEKA for medical data classification and early disease prediction. In: 2017 3rd International Conference on Computational Intelligence & Communication Technology (CICT); 2017:1–6. doi:10.1109/CIACT.2017.7977277
21. Shalev-Shwartz S, Ben-David S. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press; 2014.
22. Witten IH, Frank E, Hall MA, Pal CJ. *Data Mining, Fourth Edition: Practical Machine Learning Tools and Techniques*. 4th ed. Morgan Kaufmann Publishers Inc.; 2016.
23. Salzberg SL. C4.5: programs for machine learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993. *Mach Learn*. 1994;16:235-240. doi:10.1007/BF00993309
24. Saravanan N, Gayathri V. Performance and classification evaluation of J48 algorithm and Kendall's based J48 algorithm (KNJ48). *Int J Comput Trends Technol*. 2018;59:73-80.
25. Quinlan JR. Learning decision tree classifiers. *ACM Comput Surv (CSUR)*. 1996;28(1):71-72.
26. Succo G, Peretti G, Piazza C, et al. Open partial horizontal laryngectomies: a proposal for classification by the working committee on nomenclature of the European laryngological society. *Eur Arch Otorhinolaryngol*. 2014;271(9):2489-2496. doi:10.1007/s00405-014-3024-4
27. Amin MB, Edge S, Greene F, et al., eds. *AJCC Cancer Staging Manual (8th Edition)*. Springer International Publishing: American Joint Commission on Cancer; 2017.
28. Iocca O, Di Maio P, De Virgilio A, et al. Lymph node yield and lymph node ratio in oral cavity and oropharyngeal carcinoma: preliminary results from a prospective, multicenter, international cohort. *Oral Oncol*. 2020;107:104740. doi:10.1016/j.oraloncology.2020.104740
29. Department of Veterans Affairs Laryngeal Cancer Study Group, Wolf GT, Fisher SG, et al. Induction chemotherapy plus radiation compared with surgery plus radiation in patients with advanced laryngeal cancer. *N Engl J Med*. 1991;324(24):1685-1690. doi:10.1056/NEJM199106133242402
30. Forastiere AA, Goepfert H, Maor M, et al. Concurrent chemotherapy and radiotherapy for organ preservation in advanced laryngeal cancer. *N Engl J Med*. 2003;349(22):2091-2098. doi:10.1056/NEJMoa031317
31. Labayle J, Bismuth R. La laryngectomie totale avec reconstitution [Total laryngectomy with reconstitution]. *Ann Otolaryngol Chir Cervicofac*. 1971;88(4):219-228.
32. Bur AM, Shew M, New J. Artificial intelligence for the otolaryngologist: a state of the art review. *Otolaryngol Head Neck Surg*. 2019;160(4):603-611. doi:10.1177/0194599819827507
33. Bishop CM, Nasrabadi NM. *Pattern Recognition and Machine Learning*. Springer; 2006.
34. Obermeyer Z, Emanuel EJ. Predicting the future – big data, machine learning, and clinical medicine. *N Engl J Med*. 2016;375(13):1216-1219. doi:10.1056/NEJMp1606181
35. Zhong J, Frood R, Brown P, et al. Machine learning-based FDG PET-CT radiomics for outcome prediction in larynx and hypopharynx squamous cell carcinoma. *Clin Radiol*. 2021;76(1):78.e9-78.e17. doi:10.1016/j.crad.2020.08.030
36. McCallum A, Nigam K. A comparison of event models for naive Bayes text classification. *AAAI Conf Artif Intell*. 1998;752(1):41-48.

How to cite this article: Petruzzi G, Coden E, Iocca O, et al. Machine learning in laryngeal cancer: A pilot study to predict oncological outcomes and the role of adverse features. *Head & Neck*. 2023;45(8):2068-2078. doi:10.1002/hed.27434