

Discrimination of Vocal Folds Lesions by Multiclass Classification using Autofluorescence Spectroscopy – an ex vivo study

Olivier Gaiffe ^{1,2}, Joackim Mahdjoub ¹, Emmanuel Ramasso ³, Olivier Mauvais ¹, Thomas Lihoreau ⁴, Lionel Pazart ⁴,
Bruno Wacogne ^{3,4}, Laurent Tavernier ^{1,2}

Author affiliations

¹ Department of Otolaryngology - Head and Neck Surgery, CHU de Besançon, Besançon, France.

² Laboratoire de Nanomédecine, Imagerie et Thérapeutiques, EA4662, Université de Franche-Comté, Besançon, France.

³ Institut FEMTO-ST UMR61742, Université de Franche-Comté, ENSMM, CNRS, Besançon, France.

⁴ Inserm CIC 1431, CHU Besançon, Besançon, France.

Correspondence

Olivier Gaiffe, Department of Otolaryngology - Head and Neck Surgery, CHU Besançon, Besançon, France.

Email: ogaiffe@chu-besancon.fr

Conflict of interest disclosure

The authors declare no conflicts of interest.

Funding statement

European Union Seventh Framework Programme FP7 [288663, 2012]; the Agence Nationale de la Recherche, France [ANR-17-CE19-0005, 2017].

Data availability statement

Research data are stored in an institutional repository and will be shared upon request to the corresponding author.

Clinical trial registration

ClinicalTrial.gov Identifier: NCT03585075

Abstract

Background: Autofluorescence spectroscopy is effective for non-invasive detection but underutilized in tissue with various pathology analyses. This study evaluates whether AFS can be used to discriminate between different types of laryngeal lesions in view of assisting in vocal fold surgery and preoperative investigations.

Methods: A total of 1308 spectra were recorded from 29 vocal fold samples obtained from 23 patients. Multiclass analysis was performed on the spectral data, categorizing lesions into normal, benign, dysplastic, or carcinoma.

Results: Through an appropriate selection of spectral components and a cascading classification approach based on artificial neural networks, a classification rate of 97% was achieved for each lesion class, compared to 52% using autofluorescence intensity.

Conclusions: The ex vivo study demonstrates the effectiveness of AFS combined with multivariate analysis for accurate classification of vocal fold lesions. Comprehensive analysis of spectral data significantly improves classification accuracy, such as distinguishing malignant from precancerous or benign lesions.

Keywords

Head and Neck cancer; Vocal Cords; Autofluorescence; Spectrometry; Machine Learning;

Introduction

Early detection is a major issue in the management of laryngeal cancer, as delay can lead to a poor prognosis and or the necessity of aggressive treatments with potential severe functional impairment. Currently, histopathology remains the gold standard for the identification and extension assessment of lesions. This analysis involves collecting biopsies, generally during micro-laryngoscopy under general anesthesia, making it an inherently invasive method. The need to obtain clear surgical margins and the risk of false negatives due to incorrect sample areas can lead to an increased number of biopsies. In the case of vocal folds, repeated biopsies can lead to severe voice and swallowing disorders.

Therefore, finding minimally invasive diagnostic techniques is a major challenge. Many complementary diagnostic tools have been developed to help clinicians.¹ Because of their non-invasive nature, optical techniques are among the most promising methods.^{2,3} Current data from the literature support the idea that many cellular and subcellular changes can be detected using these methods.^{4,5}

Autofluorescence endoscopy imaging (AFEI) and autofluorescence spectroscopy (AFS) have been gaining interest during the last decades. They are accurate and fast optical methods that can be used to detect early lesions,^{6,7} and allow real-time detection of malignancies.^{8,9} These techniques are based on the measurement of tissue fluorescence emission. Under ultraviolet or blue light illumination, several endogenous fluorophores such as collagen, elastin, NADH, flavin, or porphyrin produce autofluorescence within the visible wavelength region. The shape of the fluorescence spectra depends on the excitation wavelength, the presence of fluorophores, and the optical properties of tissues. Typically, molecular changes in carcinogenesis lead to distinct fluorescence spectra.¹⁰

In the field of laryngeal oncology, many studies focused on using the intensity of the fluorescence signal (Ig) of the tissues as a biomarker to discriminate lesions.¹¹⁻¹³ Results were encouraging but showed poor specificity, as many benign pathologies caused a decrease in fluorescence intensity. A small number of studies focused on fluorescence spectroscopy,^{10,14-16} and showed differences between spectral patterns of normal and malignant tissues of the larynx. However, discriminating benign from cancerous lesions, or premalignant from malignant lesions remained impossible in these studies. Another difficulty was the way of selecting significant information contained in the spectral data. Altogether, the diagnostic benefit of AFS was low. Yet, AFEI and AFS have been successfully applied to cancer detection in several other organs, notably in the oral cavity.¹⁶⁻¹⁸ Indeed, ex vivo and in vivo, studies have shown the effectiveness of AFEI and AFS for the diagnosis of oral squamous cell carcinoma with a sensitivity of 0.82-1.00 and specificity of 0.63-1.00 depending on the series.¹⁹⁻²²

The present study's objective was to assess the accuracy of AFS in distinguishing between different types of vocal cord lesions. Our approach involved selecting a sufficient number of spectral components and employing a cascading classification method based on artificial neural networks to achieve multiclass discrimination. To evaluate the effectiveness of the method, classification models based on global fluorescence intensity or discriminant analysis were used for comparison.

Materials and Methods

Clinical Study Design

In this prospective study, we examined patients undergoing vocal folds surgery at the Besançon University Hospital between 2014 and 2016. “Fluorocord” trial received a favorable opinion from the French ethical committee (CPP Est-II) in June 2014 (NCT03585075). Informed written consent was obtained from each patient. Biopsy sites for lesions were selected on a visual examination basis by the surgeon (OM). Normal vocal fold biopsies were taken from total laryngectomy surgical specimens or cordectomy specimens for non-tumoral indication. After AFS recording, biopsies were embedded in 10% neutral formalin and sent to histopathology for diagnosis confirmation.

Fluorescence Spectroscopy Instrumentation

For each biopsy, a hyperspectral fluorescence image of $1.4 \times 1.4 \text{ mm}^2$ with a pitch of $200 \text{ }\mu\text{m}$ was obtained using the point-scan method. Two motorized translation stages were employed to move a fiber optic probe (Avantes) along the x and y axes (**Fig. 1A**). The biopsies were placed on a customized sample holder (**Fig. 1B**) and kept moist with saline. The sample was positioned in the center of the device and rested on a removable plunger that could be adjusted to accommodate different biopsy thicknesses. To ensure that the sample was flat, a $170\text{-}\mu\text{m}$ thick coverslip was inserted into the slot shown in **Fig. 1B**. The manual translation stage was then used to position the probe $100 \text{ }\mu\text{m}$ from the coverslip. The probe consisting of seven $200\text{-}\mu\text{m}$ core fibers, one for excitation and the others for collection, was used with the excitation wavelength set at 405 nm (Flexpoint). A band pass filter centered at 405 nm ($\pm 10 \text{ nm}$) was placed on the excitation path, while a low pass filter at 437 nm was used to remove the excitation light from the collection path. A spectrometer (Ocean Optics USB 4000ES) was then used to measure the fluorescence spectrum. A fluorescence spectrum recorded in the $440\text{--}750 \text{ nm}$ wavelength range, representing 1,550 different wavelengths separated by 0.20 nm , was associated with each measurement point. Spectra that saturated the sensor or had a signal-to-noise ratio below 15 dB were excluded from the study.

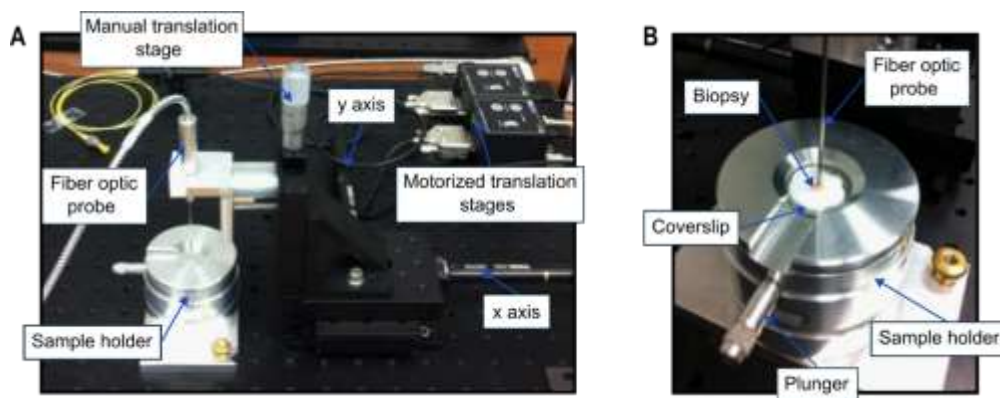


Figure 1. (A) Hyperspectral fluorescence image set-up. **(B)** Customized sample holder.

Features Extraction

The fluorescent intensity (I_g) was extracted from fluorescence spectra to be used as a reference during the classification result. This descriptor (I_g) was estimated by calculating the integral of the spectrum between 440 and 750 nm. Principal component analysis is performed directly on the centered, non-reduced raw spectra data. This method, described in detail by Jolliffe²³ and used by Eker et al.¹⁶ on autofluorescence spectra, can be considered as data or variable compression in our case. The result of PCA is a set of new dimensions, called principal components (PCs), which are uncorrelated and calculated by preserving as much variance as possible. In this case, we selected a sufficient number of PCs to retain at least 99.9% of the variance contained in the raw dataset, thus reducing the original 1,550 wavelengths to only 20 PCs.

Tissue Classification Method

The histopathological diagnosis was performed by grouping all biopsies into four categories: normal, benign, dysplasia, and carcinoma. Histopathological analyses were conducted by an independent pathologist. In this section, we propose a method for deciding on the category (referred to as a "class" hereafter) based on transmittance spectra.

The method is termed "hierarchical" because it involves a cascade of binary classifiers and can be represented using a decision tree, as illustrated in **Fig. 2**. Each level of the hierarchy is a binary classifier. The data from an unknown biopsy (test data) first goes through the first classifier ("model 1" in the figure), which provides a result that can be "normal" (class represented by "N" in the figure) or "not normal". If the result is "normal", the diagnosis is made, and the process stops. Otherwise, it means that the class could be benign (B), dysplasia (D), or carcinoma (C). Thus, the data goes through a second classifier ("model 2"). If the classifier yields the "benign" class, the diagnosis is made; otherwise, the data goes through the third classifier ("model 3"), which distinguishes between the last two classes. The utility of this cascade classification method has been demonstrated in numerous applications, including the classification of tumor and non-tumor tissues based on Raman spectroscopy.²⁴ For each spectrum (after PCA) to be tested, the hierarchical method assigns a class. However, since multiple spectra are available per biopsy, there are consequently multiple possible

classes. Therefore, we used a decision fusion method based on a majority vote, associating the biopsy with the most represented class.

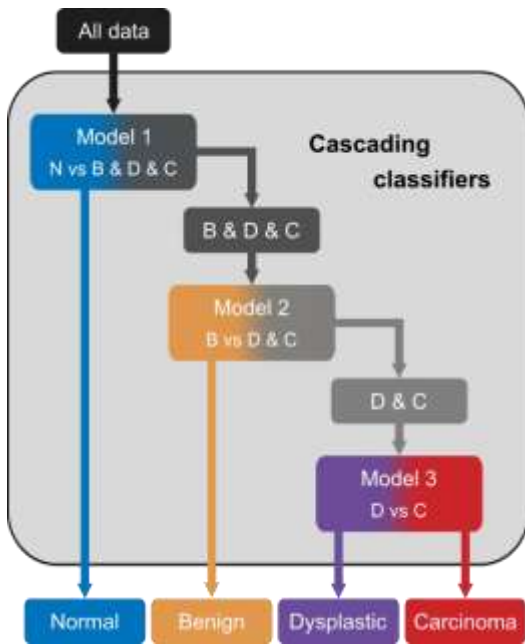


Figure 2. Cascading classification using three different models to separate 4 groups: normal (N), benign (B), dysplastic (D), and carcinoma (C).

Models 1, 2, and 3 are classifiers that take the PCA of each spectrum as input. We studied and compared two classifiers that have demonstrated their performance in many machine learning applications, namely, discriminant analysis (DA) and artificial neural networks (ANN). The network consists of a hidden layer with ten neurons. For training these two classifiers, a 5-fold cross-validation was used where the dataset was divided into 5 parts, with 4 used for training and 1 for testing. The test results provided in the following figures and tables represent the average across the 5 test sets which means that all biopsies were tested.

From evaluating the classification results, we used the following metrics: the diagnostic accuracy (Acc), the sensitivity (Se), the specificity (Sp), and the Youden index (J),²⁵ which is calculated as $J = Se + Sp - 1$, thus giving an optimal value of the discrimination threshold. Clopper-Pearson 95% confidence intervals were calculated for Se, Sp, and Acc. All data processing and statistical data analysis were performed using MATLAB™ 2019b (MathWorks Inc., Natick MA).

Results

Cohort Characteristics

A sample of 29 biopsies was used from 23 patients. For 6 patients, 2 biopsies were collected on each vocal cord.

According to the histological analysis, 5 biopsies were classified as dysplasia (including 2 moderate to severe dysplasia

and 3 carcinomas in situ (CIS)), 10 as carcinoma (including 8 Squamous Cell Carcinoma (SCC) and 2 spindle cell carcinomas), 5 as normal and 8 as benign (including 4 polyps, 2 inflammatory lesions, 1 ulceration with granuloma, 1 papilloma and 1 leukokeratosis) (**Table 1**). A total of 1308 spectra was recorded and divided as follows: 245 normal spectra, 409 benign, 227 dysplastic, and 427 malignant.

Table 1: Cohort Characteristics

Patient	Sex	Biopsy (N°)	Number of Spectra	Group	Histology
A	M	1	49	N	Normal
B	F	2	49	C	SCC
C	F	3	36	D	Moderate to severe dysplasia
		4	43	C	SCC
D	F	5	49	B	Myxoid polyp
E	M	6	49	C	SCC
F	M	7	31	C	SCC
G	M	8	49	D	CIS
H	M	9	46	B	Polyp
I	M	10	49	C	SCC
		11	49	D	Moderate dysplasia
J	M	12	49	D	CIS
		13	49	N	Normal
K	M	14	46	C	SCC
		15	49	B	Inflammation
L	M	16	46	B	Mucosal ulceration with granulation tissue
M	M	17	49	B	Inflammatory polyp
N	M	18	49	B	Leukokeratosis
O	M	19	48	B	Inflammation, hyperplasia
P	M	20	26	C	Sarcomatoid carcinoma
Q	M	21	44	D	Severe dysplasia to CIS
R	M	22	49	C	Sarcomatoid carcinoma
		23	49	N	Normal
S	M	24	31	B	Papilloma
T	M	25	36	C	SCC
U	M	26	49	C	SCC
V	M	27	42	B	Pseudo-polyp
W	M	28	49	N	Normal
		29	49	N	Normal

Abbreviations: Male (M), Female (F), Normal (N), Benign (B), Dysplastic (D), Carcinoma (C), Squamous Cell Carcinoma (SCC), and Carcinoma In Situ (CIS).

Spectral Response Patterns

The intensity I_g for each of the recorded data is represented in the box plot reported in **Fig. 3A**. The fluorescence intensity decreases from the normal to the benign group. The median value of the benign group is close to the carcinoma group. Normalized area-under-the-curve (nAUC) spectra, typical of each lesion group, are displayed in **Fig. 3B**.

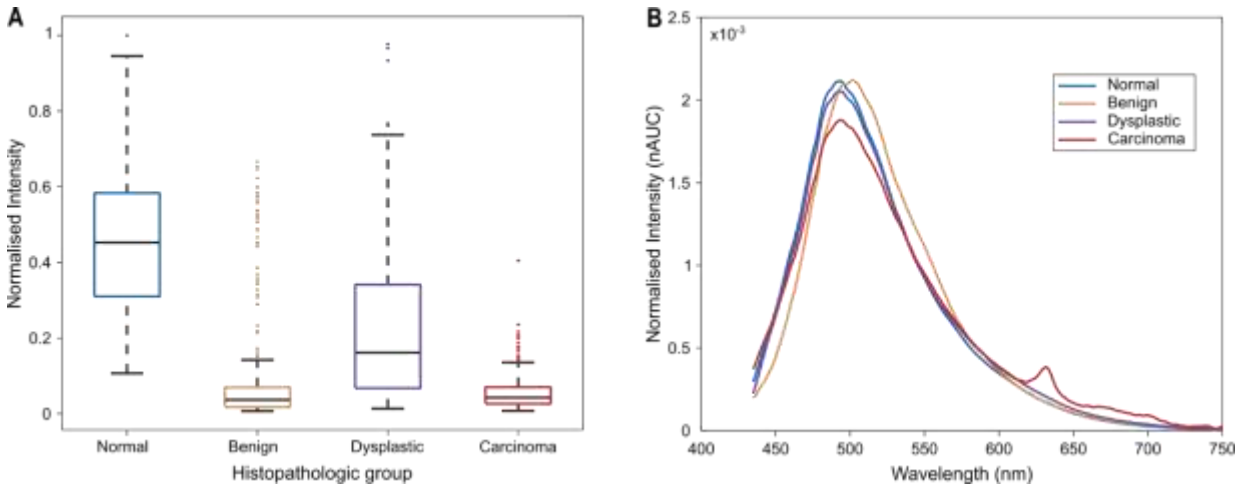


Figure 3. (A) Fluorescence intensity I_g of the four categories for all measurements normalized by the maximum of the dataset. The box boundaries indicate the first and third quartile and the black line indicates the median. The upper and lower whiskers are specified as 1.5 times the interquartile range. Dots denote outliers. **(B)** Normalized area-under-the-curve (nAUC) spectra are representative of the 4 groups: normal (biopsy 1), benign (biopsy 17), dysplastic (biopsy 12), and carcinoma (biopsy 14).

In this example, the benign case corresponded to an inflammatory polyp, and a noticeable shift in the central wavelength was observed. The dysplastic case was diagnosed as CIS. Interestingly, the corresponding spectrum did not exhibit any discernible differences compared to the normal case. On the other hand, the spectrum obtained from an invasive squamous cell carcinoma showed a distinct and highly differentiated spectral shape, characterized by a prominent peak at 632 nm. However, it is important to note that this specific spectral feature, although easily identifiable, was not systematically observed in all cancer biopsies included in our study, representing only 18% of this class of spectra.

Classification of Models

The hierarchical classification method detailed in the "Tissue classification methods" section consists of 3 binary classifiers (models 1, 2, 3 in Fig. 2). We considered two classification methods: either an ANN or a DA, and compared thereafter. In this section, we consider the classifiers individually, without considering the hierarchy. The goal is to

determine the number of principal components for each of them. The hierarchical method will then be implemented and evaluated in the following section with the number of components determined in this section.

For each classifier, cross-validation is used to estimate the number of principal components (PC) of PCA to maximize performance in validation. To achieve this, we vary the number of components from 1 to 20. For each value, 80% of the training set is used to estimate the parameters of a given classifier, and 20% is used to estimate performance in validation. This process is repeated 5 times, and the average validation performance is used to choose the most suitable number of components, which is then used to evaluate the classifier on the testing data.

Evaluation of model 1 – Normal (N) vs not normal: The objective of this first model was to distinguish healthy from diseased tissues. Results are presented in **Table 2**. The classification results based on Ig intensity were equivalent regardless of the classification method (ANN or DA) used. Metrics Se, Sp, and Acc ranged from 0.87-0.89, 0.80-0.84, and 0.87, respectively. J was slightly higher for ANN compared to DA with values of 0.71 and 0.69 respectively.

Fig. 4A shows the evolution of the J-index as a function of the number of PCs used for DA and ANN. We observe that ANN remained higher than DA for all numbers of PCs. In the case of ANN, the J-index rapidly increased up to 5 PCs and then reached a plateau, in contrast with the classification by DA which required 8 PCs, and therefore more information, to stabilize. The fluorescence spectrum provided a significant improvement leading to an increase in the J-index of 12% (DA) and 13% (ANN). Classification results obtained from spectral data were superior to the results obtained using Ig on all metrics. The maximum value was 0.97 (CI 0.97-1.00), obtained using ANN with a sensitivity of 0.99 (CI 0.99-1.00) and a specificity of 0.98 (CI 0.95-0.99).

Table 2: Classification results of sub-models

Classifier		Mean (CI 95%)			
		Sensitivity	Specificity	Accuracy	J-index
N vs (B, D & C) “model 1”					
Intensity	DA	0.89 (0.87-0.91)	0.80 (0.74-0.85)	0.87 (0.85-0.89)	0.69 (0.61-0.76)
	ANN	0.87 (0.85-0.89)	0.84 (0.79-0.88)	0.86 (0.84-0.88)	0.71 (0.64-0.77)
Spectra	DA	0.92 (0.90-0.94)	1.00 (0.99-1.00)	0.94 (0.92-0.95)	0.92 (0.89-0.94)
	ANN	0.99 (0.98-0.99)	0.98 (0.95-0.99)	0.99 (0.98-0.99)	0.97 (0.93-0.98)
B vs (D & C) “model 2”					
Intensity	DA	0.56 (0.52-0.60)	0.86 (0.82-0.89)	0.68 (0.65-0.70)	0.42 (0.34-0.49)
	ANN	0.85 (0.82-0.88)	0.44 (0.39-0.49)	0.69 (0.66-0.72)	0.29 (0.21-0.37)
Spectra	DA	0.60 (0.56-0.64)	0.99 (0.98-1.00)	0.75 (0.72-0.78)	0.59 (0.54-0.64)
	ANN	0.86 (0.83-0.89)	0.86 (0.82-0.89)	0.86 (0.84-0.88)	0.72 (0.65-0.78)
D vs C “model 3”					

Intensity	DA	0.96 (0.94-0.98)	0.53 (0.46-0.60)	0.81 (0.78-0.84)	0.49 (0.40-0.58)
	ANN	0.95 (0.93-0.97)	0.55 (0.48-0.62)	0.81 (0.78-0.84)	0.50 (0.41-0.59)
Spectra	DA	0.95 (0.93-0.97)	0.78 (0.72-0.83)	0.89 (0.87-0.91)	0.73 (0.65-0.80)
	ANN	0.97 (0.95-0.98)	0.92 (0.85-0.95)	0.95 (0.93-0.97)	0.89 (0.83-0.93)

Abbreviations: Normal (N), Benign (B), Dysplastic (D), Carcinoma (C), Discriminant Analysis (DA), Artificial Neural Network (ANN), Accuracy (Acc), Sensitivity (Se), Specificity (Sp), Youden index (J).

Evaluation of model 2 – Benign (B) vs not benign: The second model aimed at separating benign tissue from precancerous and cancerous tissue. The results obtained using Ig (**Table 2**) feature Se and Sp values near 0.50. DA gave the largest J-index with a value of 0.42 (CI 0.34-0.49) which was 14% higher compared to ANN. The estimated Se and Sp were 0.56 (CI 0.52-0.60) and 0.86 (CI 0.82-0.89). For the spectral data, the trend was similar to the first model with a rapid increase up to 5 PCs followed by a plateau (**Fig. 4B**). For all PCs, the results obtained by ANN were superior to DA. The maximum value of the J-index was 0.72 (CI 0.65-0.78). The corresponding Se and Sp parameters were 0.86 (CI 0.83-0.89) and 0.86 (CI 0.82-0.89), respectively. The fluorescence spectral information made it possible to obtain a classification gain on the J-index of 15% compared to intensity. The classification improved from a J-index of 0.42 for Ig + DA to 0.72 for PCA+ANN with the addition of spectral information.

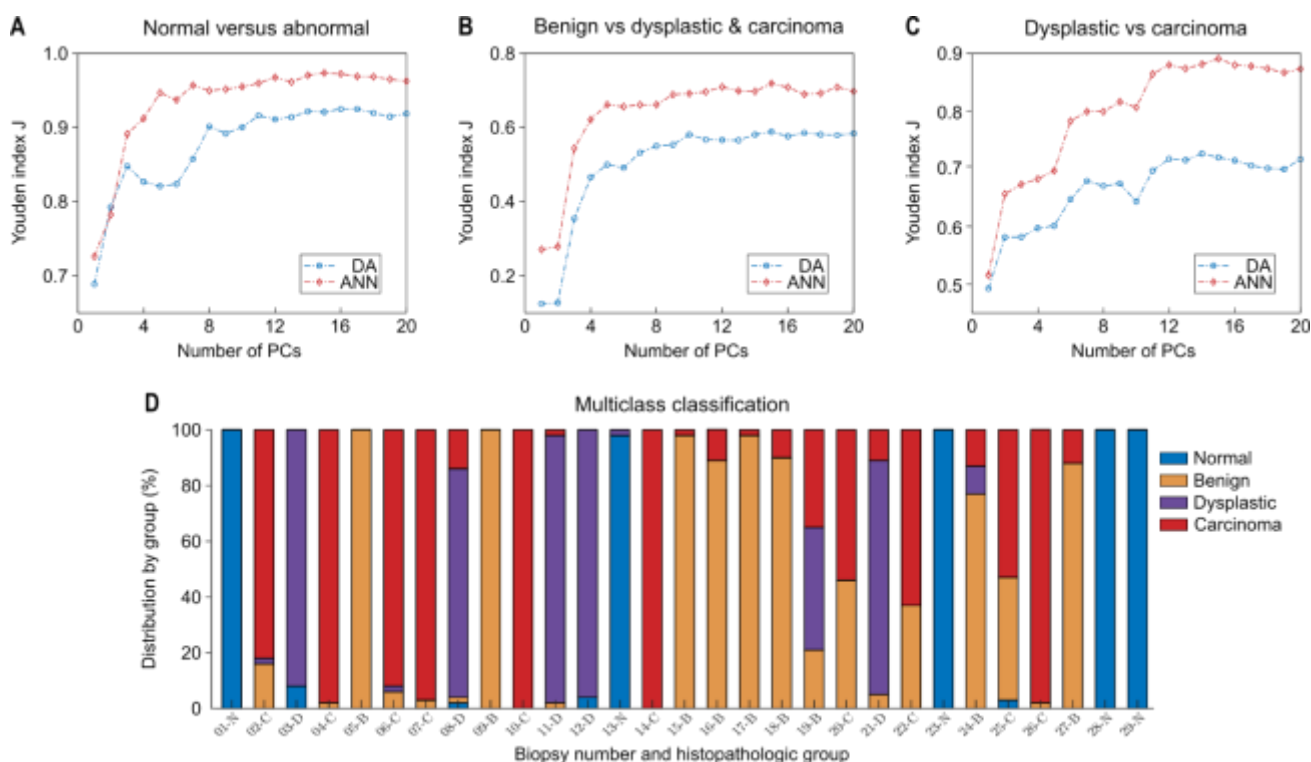


Figure 4. (A - C) Youden index (J) according to the number of principal components for DA and ANN classifiers. (D) Multiclass classification results by cascading classifiers represented by the percentage of membership per class of each biopsy.

Evaluation of model 3 – Dysplastic (D) vs Carcinoma (C): The last model sought to discriminate dysplastic tissue from malignant tissue. The classification results using Ig were similar regardless of the classification method (**Table 2**). The J-index was between 0.49 and 0.50 for Se and Sp varying from 0.94 to 0.96 and from 0.53 to 0.55, respectively. The increase in the J-index (**Fig. 4C**) with the number of PCs showed three distinct plateaus when the spectral information was included. The model required a larger number of PCs and therefore more spectral information (11 PCs) to reach the last plateau. The J-index by ANN was higher with a value of 0.89 (IC 0.83-0.93) against 0.73 (IC 0.65-0.80) by DA. The Se and Sp parameters were 0.97 (CI 0.95-0.98) and 0.92 (CI 0.88-0.95), respectively. As in the previous cases, using spectral information increased the classification efficiency by 20% compared with Ig.

In conclusion, the evaluation of classifiers revealed that the ANN is the most effective classification method, requiring 5 PCs for the first two stages of the hierarchy and 11 PCs for the third stage.

Multiclass Classification

The hierarchical classification method (**Fig. 2**) is evaluated in this section using the ANN with the number of components determined in the previous section, namely, 5 PCs for the first two stages of the hierarchy and 11 PCs for the third stage.

When taking each spectrum independently and applying the majority voting rule, the accuracy was 89%. 57% of the errors were related specifically to 4 biopsies diagnosed as inflammation - hyperplasia, SCC, and two spindle-cell carcinomas. The misclassification between the carcinoma and benign classes alone accounts for 66% of the errors.

For each biopsy, a membership percentage by class was obtained (**Fig. 4D**). Considering a single class per biopsy, defined by the predominant category contained in the analyzed scan, leads to a result of 28 out of 29 well-classified biopsies, i.e. to an accuracy of 0.97 versus 0.52 with fluorescence intensity alone.

Discussion

In this prospective trial, we demonstrated that autofluorescence spectroscopy (AFS) is an efficient way of discriminating vocal folds lesions into 4 classes. The use of all the information contained in fluorescence spectra combined with multivariate analysis provided good classification performance, even in situations where distinction could not be made in the literature¹⁵ (e.g. malignant vs. premalignant or benign lesions). Overall, 28 out of the 29 lesions were correctly recognized by our multiclass model.

To compare our classification results based on the analysis of fluorescence intensity (Ig) only, we extracted the Youden index from data taken from the literature.^{11,13} Youden indices ranging from 0.76 to 0.90 were reported, higher than the 0.71 obtained in the present case. Indeed, in these studies, the authors used the AFEI to select the area to be biopsied. The

clinicians, therefore, had additional information such as the structural modifications of the lesions, the tissue architecture, or the neovascularization.

Only 3 studies evaluated AFS on laryngeal pathologies to separate healthy tissue from diseased tissue, yet excluding benign lesions from the analyses. The conclusions were that AFS could not, or barely, improve the diagnosis compared to AFEI. We hypothesize that the recorded spectral data were not fully exploited in these works. Arens et al.¹⁰ only examined average spectra maxima, which amounts to exploiting Ig, while Rydell et al.¹⁵ compared intensity ratios at different wavelengths. Only Eker et al.¹⁶ used several multivariate analysis methods, including PCA with 4 PCs (4 PCs + QDA), and found a J-index of 0.81. By applying the same restrictions to our classification parameters, we obtained a similar result of 0.82. Increasing the J-index required analyzing more PCs: comparatively, we needed at least 8 PCs to obtain 0.90 with QDA and at least 7 PCs with ANN to obtain a J greater than 0.95.

Regarding multivariate analyses, PCA+ANN proved to be more effective than PCA+DA, particularly in isolating benign cases and in separating dysplasia from carcinoma. These results support previous studies in the field of oral oncology: Van Staveren et al.²⁷ obtained Se of 0.86 and Sp of 1.00 when isolating leukoplakia with ANN. Nayak et al.⁸ also found excellent diagnostic performances using AFS to discriminate between normal, precancerous, and malignant lesions of the oral cavity (excluding benign lesions), with superiority of ANN.

In our study, model 2 isolating the benign cases had the weakest classification result. The inclusion of benign samples leads to classification difficulties, as fluorescence characteristics are very similar to those of malignant lesions.^{28,29} This may be due to the absorption of the excitation light by hemoglobin and inflammatory cells.^{17,39} Another difficulty was observed in the literature to discriminate between dysplastic and malignant lesions either with AFEI or with AFS.^{15,31} In model 3, a distinction can be made using PCA+ANN with 15 PCs. To our knowledge, this model is the first to successfully discriminate between dysplasia or CIS and invasive cancer for vocal cord tissue. This supports the idea that using more spectral information leads to better classification results.

Concerning the multiclass classification (**Fig. 4D**), a single biopsy was incorrectly classified as dysplastic when it was defined as benign by the pathologist. The classification result of the sample was dispersed into 3 classes, with 21% of information classified as benign, 44% as dysplastic, and 35% as carcinoma. The histology of the sample concluded with very hyperkeratotic inflammatory tissue. This is consistent with previous studies^{28,30} that found SCC false negatives in AFEI related to extreme hyperkeratosis. Indeed, hyperkeratosis leads to poor penetration of light irradiation, therefore reducing fluorescence emission from the underlying cell layers and thus mimicking a dysplastic or a malignant lesion.

Moreover, this patient had many local recurrences during his follow-up, and a verrucous carcinoma was eventually diagnosed. This entity is a rare and low-grade variant of SCC, which frequently leads to diagnosis difficulties.

Three correctly classified carcinoma samples showed between 37% and 46% of their distributed spectra as benign. Two of them were spindle-cell SCC: rare, aggressive, and poorly differentiated variants of SCC which are often diagnostic challenges for pathologists. Studies showed that this type of tumor could often mimic benign mesenchymal proliferation.^{32,33} We hypothesize that these histopathologic features may influence fluorescence findings as well. The third case was a micro-invasive SCC with underlying inflammatory stroma, which may account for our results.

The main limit of our study was the number of samples per class. The monocentric nature of the study also limits its generalization at this stage. A multicenter trial would therefore allow increasing the number of samples and expanding the population. In addition, this ex vivo study only partially represents the clinical reality as it limits the presence of blood that could disturb the sample measurements in an actual surgical context. It will thus be necessary to confirm this model in vivo.

Autofluorescence spectroscopy combined with multivariate analysis has proven effective in classifying vocal cord lesions. AFS was shown to increase classification accuracy into 4 major categories up to 97%, versus 52% when considering fluorescence intensity alone. A comprehensive analysis of the principal components allowed us to distinguish lesion types, even in difficult situations. In particular, discriminating malignant from precancerous or benign lesions was successfully achieved. In the future, AFS may be used to manage laryngeal pathology as an accurate and minimally invasive diagnostic tool when combined with clinical examination. Specifically, the multispectral mapping technique could aid in the selection of a more representative and severe biopsy site for larger lesions, as well as in the determination of surgical resection margins for malignant lesions.

References

- ¹ Mattos LS, Acemoglu A, Geraldles A, Laborai A, Schoob A, Tamadazte B, et al. μ RALP and Beyond: Micro-Technologies and Systems for Robot-Assisted Endoscopic Laser Microsurgery. *Front Robot AI*. 2021;8:1-19.
- ² Wong BJF, Ilgner J, editors. *Biomedical Optics in Otorhinolaryngology*. New York: Springer; 2016.
- ³ Etievant A, Monnin J, Lihoreau T, Tamadazte B, Rougeot P, Magnin E, et al. Comparison of Noninvasive Imagery Methods to Observe Healthy and Degenerated Olfactory Epithelium in Mice for the Early Diagnosis of Neurodegenerative Diseases. *Front Neuroanat*. 2020;14:34.
- ⁴ van Schaik JE, Halmos GB, Witjes MJH, Plaat BEC. An overview of the current clinical status of optical imaging in head and neck cancer with a focus on Narrow Band imaging and fluorescence optical imaging. *Oral Oncol*. 2021;121:105504.
- ⁵ Girerd C, Lihoreau T, Rabenoroosa K, Tamadazte B, Benassarou M, Tavernier L, et al. In Vivo Inspection of the Olfactory Epithelium: Feasibility of Robotized Optical Biopsy. *Ann Biomed Eng*. 2018;46(11):1951-61.
- ⁶ Dhingra JK, Perrault DF, McMillan K, Rebeiz EE, Kabani S, Manoharan R, et al. Early Diagnosis of Upper Aerodigestive Tract Cancer by Autofluorescence. *Arch Otolaryngol Head Neck Surg*. 1996;122(11):1181-6.
- ⁷ Omar E. Current concepts and future of noninvasive procedures for diagnosing oral squamous cell carcinoma - a systematic review. *Head Face Med*. 2015;11(1):6.
- ⁸ Nayak GS, Kamath S, Pai KM, Sarkar A, Ray S, Kurien J, et al. Principal component analysis and artificial neural network analysis of oral tissue fluorescence spectra: Classification of normal premalignant and malignant pathological conditions. *Biopolymers*. 2006;82(2):152-66.
- ⁹ Nazarian S, Gkouzionis I, Kawka M, Jamroziak M, Lloyd J, Darzi A, et al. Real-time Tracking and Classification of Tumor and Nontumor Tissue in Upper Gastrointestinal Cancers Using Diffuse Reflectance Spectroscopy for Resection Margin Assessment. *JAMA Surg*. 2022;157(11):e223899.
- ¹⁰ Arens C, Reußner D, Neubacher H, Woenckhaus J, Glanz H. Spectrometric measurement in laryngeal cancer. *Arch Otorhinolaryngol*. 2006;263(11):1001-7.
- ¹¹ Ge MW, Ni HT, Huang JW, Fan ZH, Shen WQ, Chen HL. Diagnostic value of autofluorescence laryngoscope in early laryngeal carcinoma and precancerous lesions: A systematic review and meta-analysis. *Photodiagnosis Photodyn Ther*. 2021;35:102460.
- ¹² Wu C, Gleysteen J, Teraphongphom NT, Li Y, Rosenthal E. In-vivo optical imaging in head and neck oncology: basic principles, clinical applications and future directions. *Int J Oral Sci*. 2018;10(2):10.

- 13 Mannelli G, Cecconi L, Gallo O. Laryngeal preneoplastic lesions and cancer: challenging diagnosis. Qualitative literature review and meta-analysis. *Crit Rev Oncol Hematol*. 2016;106:64-90.
- 14 Winiarski P, Szewczyk-Golec K, Orłowski P, Kałużna E, Wamka M, Mackiewicz-Nartowicz H, et al. Autofluorescence spectroscopy in the differentiation of laryngeal epithelial lesions – preliminary results. *Acta Otolaryngol*. 2016;136(6):580-4.
- 15 Rydell R, Eker C, Andersson-Engels S, Krogdahl A, Wahlberg P, Svanberg K. Fluorescence investigations to classify malignant laryngeal lesions in vivo. *Head Neck*. 2008;30(4):419-26.
- 16 Eker C, Rydell R, Svanberg K, Andersson-Engels S. Multivariate analysis of laryngeal fluorescence spectra recorded in vivo. *Lasers Surg Med*. 2001;28(3):259-66.
- 17 Gillenwater A, Jacob R, Ganeshappa R, Kemp B, El-Naggar AK, Lynn Palmer J, et al. Noninvasive Diagnosis of Oral Neoplasia Based on Fluorescence Spectroscopy and Native Tissue Autofluorescence. *Arch Otolaryngol Head Neck Surg*. 1998;124(11):1251.
- 18 Poh CF, Zhang L, Anderson DW, Durham JS, Williams RM, Priddy RW, et al. Fluorescence visualization detection of field alterations in tumor margins of oral cancer patients. *Clinical Cancer Research*. 2006; 12(22):6716-22.
- 19 De Veld DCG, Skurichina M, Witjes MJH, Duin RPW, Sterenborg DJCM, Roodenburg JLN. Autofluorescence and diffuse reflectance spectroscopy for oral oncology. *Lasers Surg Med*. 2005;36(5):356-64.
- 20 Poh CF, Anderson DW, Durham JS, Chen J, Berean KW, MacAulay CE, et al. Fluorescence Visualization–Guided Surgery for Early-Stage Oral Cancer. *JAMA Otolaryngol Head Neck Surg*. 2016;142(3):209-16.
- 21 Luo X, Xu H, He M, Han Q, Wang H, Sun C, et al. Accuracy of autofluorescence in diagnosing oral squamous cell carcinoma and oral potentially malignant disorders: a comparative study with aero-digestive lesions. *Sci Rep*. 2016;6(1):29943.
- 22 Durham JS, Brasher P, Anderson DW, Yoo J, Hart R, Dort JC, et al. Effect of Fluorescence Visualization–Guided Surgery on Local Recurrence of Oral Squamous Cell Carcinoma. *JAMA Otolaryngol Head Neck Surg*. 2020;146(12):1149.
- 23 Jolliffe IT. *Principal Component Analysis*. 2nd ed. New York: Springer; 2002.
- 24 Cals FLJ, Koljenović S, Hardillo JA, Baatenburg de Jong RJ, Bakker Schut TC, Puppels GJ. Development and validation of Raman spectroscopic classification models to discriminate tongue squamous cell carcinoma from non-tumorous tissue. *Oral Oncol*. 2016;60:41-47.
- 25 Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. 2nd ed. New York: Springer; 2009.
- 26 Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3(1):32-5.

- 27 Van Staveren HJ, Van Veen RL, Speelman OC, Witjes MJH, Star WM, Roodenburg JLN. Classification of clinical autofluorescence spectra of oral leukoplakia using an artificial neural network: a pilot study. *Oral Oncol.* 2000;36(3):286-93.
- 28 Arens C, Dreyer T, Glanz H, Malzahn K. Indirect autofluorescence laryngoscopy in the diagnosis of laryngeal cancer and its precursor lesions. *Arch Otorhinolaryngol.* 2004;261(2):71-76.
- 29 Paczona R, Temam S, Janot F, Marandas P, Luboinski B. Autofluorescence videoendoscopy for photodiagnosis of head and neck squamous cell carcinoma. *Arch Otorhinolaryngol.* 2003;260(10):544-8.
- 30 Malzahn K, Dreyer T, Glanz H, Arens C. Autofluorescence Endoscopy in the Diagnosis of Early Laryngeal Cancer and Its Precursor Lesions. *Laryngoscope.* 2002;112(3):488-93.
- 31 Saetti R, Derosas F, Silvestrini M, Narne S. Efficacy of autofluorescence videoendoscopy in the diagnosis of laryngeal lesions. *Acta Otorhinolaryngol Ital.* 2007;27(4):181-5.
- 32 Neha B, Shashi D, Seema R. Spindle Cell Squamous Cell Carcinoma of Head and Neck Region: a Clinicopathological and Immunohistochemical Study. *Indian J Surg Oncol.* 2021;12(4):699-705.
- 33 Prieto-Granada CN, Xu B, Alzumaili B, Al Rasheed MRH, Eskander A, Enepekides D, et al. Clinicopathologic features and outcome of head and neck mucosal spindle cell squamous cell carcinoma. *Virchows Arch.* 2021;479(4):729-39.