DETECTION OF *IN VIVO* ORAL EPITHELIAL CANCER USING FLUORESCENCE LIFETIME IMAGING

A Thesis

by

DAE YON HWANG

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Jim Xiuquan Ji
Co-Chair of Committee,	Javier A. Jo
Committee Members,	Ulisses Braga-Neto
	Sebastian Hoyos
Head of Department,	Miroslav M. Begovic

May 2016

Major Subject: Electrical Engineering

ABSTRACT

Endogenous fluorescence lifetime imaging microscopy (FLIM) provides a nondestructive means to interrogate the biochemical composition of biological tissues. Therefore, it has the potential to identify tissue pre-malignant and malignant transformation. In this study, we evaluate the potential of endogenous FLIM for detecting benign oral lesions from pre-malignant and malignant oral lesions. Using a database of FLIM images (n=20) obtained *in vivo* from the oral cavity of patients undergoing tissue biopsy, we were able to identify specific features from the characteristic FLIM signal of benign, mild dysplastic, and cancerous oral lesions. These features were used to train statistical classification rules aimed to detect benign lesions from either dysplastic and cancerous lesions. Our results indicated that dysplastic and cancerous lesions could be detected from benign lesions with sensitivity of ~89% and specificity of ~95%. Our future efforts are focused on further developing our classification algorithms with additional FLIM *in vivo* data.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Jo who gave me a chance to do research and motivated me to complete this work. I would also like to appreciate my committee members, Dr. Ji, Dr. Braga-Neto, and Dr. Hoyos, for their guidance and support throughout the course of this research.

Thanks to my friends and colleagues for making my time at Texas A&M University a great experience.

Finally, special thanks to my family and girlfriend for their encouragement and support during the graduate school.

TABLE OF CONTENTS

Page

ABSTRACT ii

ACKNOWLEDGEMENTS iii

TABLE OF CONTENTS iv

LIST OF FIGURES v

LIST OF TABLES vi

1. INTRODUCTION 1

2. METHOD 6

2.1 Multispectral FLIM endoscope [19] 6
2.2 *In vivo* imaging 8
2.3 FLIM data processing 9
2.4 Feature pre-selection 12
2.5 Classification problems 15
2.6 Feature selection 16
2.7 Classification algorithms 16
2.8 Performance quantification 19

3. RESULT 20

3.1 Histopathology evaluation 20
3.2 Classification – Gingiva 21
3.3 Classification – Tongue 24
3.4 Classification – Gingiva and Tongue 27

4. CONCLUSIONS 31

REFERENCES 32

LIST OF FIGURES

FIGURE		Page
1	Photograph and schematic of the FLIM endoscopy system	7
2	*In vivo* imaging of a dysplasia lesion	8
3	Composition of FLIM data	9
4	Enlarged actual signals of a pixel in FLIM data	10
5	Deconvolved fluorescence decays	12
6	Final results of FLIM data processing	12
7	Median plots from gingiva and tongue.	14
8	Scatter plot based on gingiva training datasets	22
9	Classification maps of gingiva	23
10	Scatter plot based on tongue training datasets	25
11	Classification maps of tongue	26
12	Scatter plot based on total training datasets	28
13	Classification maps of gingiva and tongue	29

LIST OF TABLES

TABLE		Page
1	Summary of histopathology distribution	20
2	LOO sample-level confusion matrix of gingiva area	22
3	LOO pixel-level confusion matrix of gingiva area	24
4	LOO sample-level confusion matrix of tongue area	25
5	LOO pixel-level confusion matrix of tongue area	27
6	LOO sample-level confusion matrix of entire area	28
7	LOO pixel-level confusion matrix of entire area	30

1. INTRODUCTION

The autofluorescence imaging and spectroscopy can probe alternations noninvasively that occur during malignant process [1]. Fluorescence that comes from epithelial tissue is resulted from multiple fluorophores and affected by scattering and absorption as light propagates through the stroma and epithelium [2-4]. Reduced form of nicotinamide adenine dinucleotide (NADH) and mitochondrial metabolic coenzymes flavin adenine dinucleotide (FAD) have been known as the major sources of fluorescence in the epithelial layer. Collagen cross-links are the main source of autofluorescence in the underlying stroma. In normal oral epithelial tissue, stromal collagen is the major source of fluorescence, although there is also epithelial fluorescence. As dysplasia and malignancy developing, epithelial NADH/FAD fluorescence increases, while collagen fluorescence decreases. On the other hand, benign inflammation appears a decrease in both stromal collagen and epithelial NADH/FAD fluorescence. Bright fluorescence can also be found from the keratinized, superficial epithelial layer, which is often shown in normal oral tissue from specific anatomic sites such as the palate and the gingiva, as well as in clinically apparent leukoplakia. Furthermore, endogenous porphyrin fluorescence has been related to oral epithelial malignancy, although its accumulation is considered to be the result of microbial synthesis [5, 6].

Steady-state wide-field imaging and fluorescence point-spectroscopy have been researched for the clinical diagnosis of oral epithelial cancer. In point-spectroscopy, a

localized area of the tissue is explained with a single or multiple excitation wavelengths and the fluorescence emission is resolved over a broad spectral range. Novel fiber-probe designs have been developed for depth-sensitive fluorescence spectroscopy. Although it has a relatively rich signal, detection of benign from (pre)malignant oral lesions through fluorescence spectroscopy was not fully researched. Moreover, because of its interrogation area and limited spatial resolution, it is impractical for screening large areas of the oral mucosa. In wide-field imaging, the tissue is explained with a single excitation wavelength and the fluorescence emission is pictured over a single broad emission band. Thanks to its capability for real-time screening large areas, wide-field fluorescence imaging is already being applied on commercially and clinically available. Unfortunately, several studies have contradicted its utility as an adjunct tool for early diagnosis and screening because of its reported low specificity. There are some key factors that explain its lack of specificity: Subjectivity in its interpretation and limited characterization of the fluorescence. Objective interpretation based on image processing and multispectral imaging has been recently argued to address these problems, although we need to assess its utility [7, 8]. Many other reasons may explain why various steady-state fluorescence imaging/spectroscopy studies have been so far inconclusive. For example, since steady-state fluorescence imaging/spectroscopy are based on relative or absolute fluorescence intensity measurements, they are sensitive to experimental artifacts which are hard to control during clinical interventions, and other factors not associated with dysplastic process (i.e. hyperplasia, keratinization, and blood optical interference) [9-11].

On the other hand, fluorescence lifetime measurements are less sensitive to such artifacts which means more robust for clinical applications [12]. In addition, since most endogenous fluorophores have broad emission bands, tissue characterization based on only spectral/intensity changes may be hard to get sufficient accuracy for clinical detection. However, fluorescence lifetime can resolve fluorophores with heavily overlapping spectra [13]. In particular, fluorescence from NADH and collagen has overlapping emission spectra and same excitation maximum, but very distinct fluorescence lifetimes. Furthermore, fluorescence lifetime is sensitive to alternations in cellular physiology [13]. In particular, the relative concentration of the free and bound forms of both FAD and NADH are associated with the metabolic state of the tissue; thus, providing additional means for detecting highly active precancerous tissue. While the free and bound forms of NADH have virtually the same absorption and emission spectra (the same applies for FAD), they show very distinct fluorescence lifetimes [14]. Thus, differences in tissue fluorescence resulting from compositional, morphological and functional transformations related to (pre)malignancy would be more accurately quantified based on fluorescence lifetime measurements.

In spite of its significant advantages, fluorescence lifetimes measurements have not been extensively evaluated for the detection of oral (pre)malignancy. Wang *et al.* performed *in vivo* autofluorescence TRFS measurements at ~630 nm upon 408 nm excitation in patients with suspicious (pre)malignant lesions and reported that dysplastic lesions could be found from benign lesions (both epithelial hyperplasia and verrucous) with 75%

specificity and 93% sensitivity on a small database of 38 lesions [15]. Marcu *et al.* performed *in vivo* fluorescence TRFS measurements in the spectral range of 360-650 nm upon 337 nm excitation in the hamster cheek-pouch model and found three fluorescence emission peaks with distinct lifetimes at ~390 nm, 460 nm and 635 nm, which were related to collagen, NADH and porphryrin. In addition, they reported that normal epithelium could be identified from (pre)malignant lesions with 100% specificity and sensitivity, though their database did not include benign lesions. Even if these two pilot studies have been encouraging, point-spectroscopy is impractical for screening large areas of the oral mucosa. Skala *et al.* experimented endogenous multiphoton FLIM in a hamster model of oral cancer [16]. Their results showed a decrease in bound NADH lifetimes with dysplasia consistent with neoplastic metabolism. In a subsequent study, they reported an increase in bound FAD lifetime with dysplasia, reflecting a reduction in intracellular NAD+ expected with malignancy [17]. Though multiphoton FLIM microscopy gives extremely high spatial resolution both axially and laterally, it has very limited field of view and intrinsically slow acquisition speed. Marcu *et al.* reported a flexible endoscope that was used to collect 26 endogenous FLIM images from tumors and surrounding normal tissue in 10 patients undergoing head and neck surgery [18]. They observed that tumor areas appeared significantly shorter fluorescence lifetime values and weaker fluorescence intensity, compared to normal areas; however, histopathology was not used as the gold standard and the acquisition time of their instrument was too long.

The ultimate goal of this research is to validate and develop a novel imaging clinical tool that acquires and processes multispectral FLIM images of the oral mucosa *in vivo* and in real time for noninvasive real-time detection of precancerous and cancerous lesions. One key specific aim of this research is to develop algorithms for automated *in vivo* detection of oral precancer and cancer based on FLIM imaging, which was the focus of this thesis. This work represents the first demonstration of endogenous FLIM endoscopic imaging for automated detection of a wide range of oral benign lesions (including inflammation) from oral pre-cancerous (mild-dysplasia) and cancerous (squamous cell carcinoma, SCC) lesions.

2. METHOD

2.1 Multispectral FLIM endoscope [19]

The system consisted of a handheld box fitted with a custom-designed rigid endoscope, as shown in Figure 1 (a). The schematic of the FLIM endoscopy system is shown in Figure 1 (b). A frequency-tripled Q-switched Nd:YAG laser is applied as the excitation source. The excitation fiber core diameter and the objective lens focal length determine the lateral resolution (~100 um) and the field of view (FOV=11 mm diameter) of the system. The fluorescence emission is separated by a dichroic mirror (DM1) and then collected by a multimode fiber (200 μm core diameter), which delivers the emission to a multispectral detection unit outside of the handheld box. In this unit, a set of filters (F1-F3) and dichroic mirrors (DM2, DM3) separate the emission into multiple spectral bands, each one coupled into separate multimode fibers of different lengths that give an optical delay between each spectral band. Therefore, for a single excitation pulse, a single detector can be used for recording multiple decays corresponding to different spectral bands. The spectral bands can be divided based on the targeted fluorophores. We chose the 390±20 nm, 452±22.5 nm, and >500 nm bands to distinguish emission from three tissue endogenous fluorophores: collagen, NADH, and FAD, respectively. The multispectral fluorescence signal is observed by a multichannel plate photomultiplier tube (MCP-PMT, 25 ps TTS), followed by a preamplifier before being digitized at 6.25 GS/s. The multispectral FLIM data is composed of three fluorescence decays per pixel (one per emission band). After temporal deconvolution of the instrument response from

the measured fluorescence decay, nine images were made to quantify the fluorescence emission of the samples (Figure 2): absolute integrated fluorescence intensity (I1, I2 and I3), normalized integrated fluorescence intensity $(\frac{I_1}{(I_1+I_2+I_3)}, \frac{I_2}{(I_1+I_2+I_3)}, \frac{I_3}{(I_1+I_2+I_3)})$ and average lifetime maps for each of the three emission bands. The average lifetime was calculated via $\tau_{ave} = \frac{\sum th(t)}{\sum h(t)}$, where h(t) is the deconvolved fluorescence temporal decay at a given pixel, and t is the time vector.

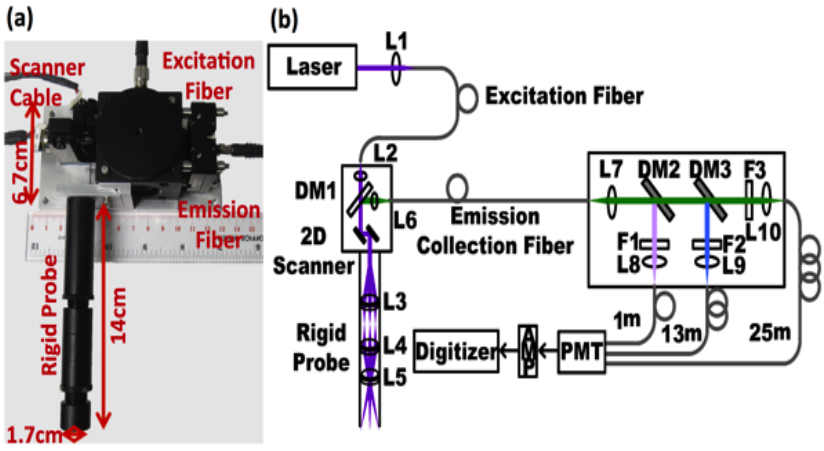

Figure 1 Photograph and schematic of the FLIM endoscopy system. (a) Photograph of the handheld rigid endoscope; (b) Schematic of the multispectral FLIM rigid endoscope system. DM: Dichroic mirror, L: Lens, F: Filter.

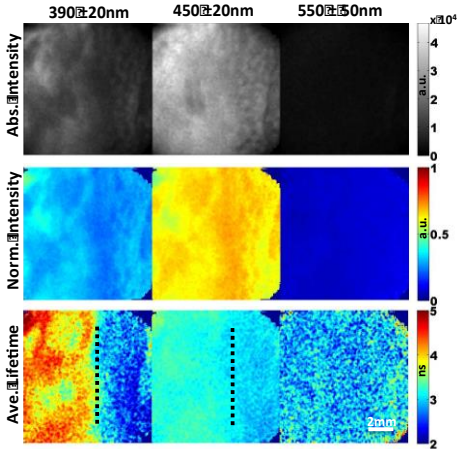

Figure 2 *In vivo* imaging of a dysplasia lesion. (a) Absolute fluorescence intensity maps; (b) Normalized fluorescence intensity maps; (c) Fluorescence lifetime maps.

2.2 *In vivo* imaging

A total of 20 patients undergoing tissue biopsy of suspicious lesions were recruited. The imaging protocol was approved by the Institutional Review Board at Texas A&M University and Baylor College of Dentistry. Prior to biopsy sample resection, the physician took a multispectral FLIM image from the clinical oral lesion with our endoscope. An additional multispectral FLIM image was acquired from a normal area on the contralateral side. The acquisition time was less than 2 seconds per image. The biopsy samples were then resected following standard procedures.

2.3 FLIM data processing

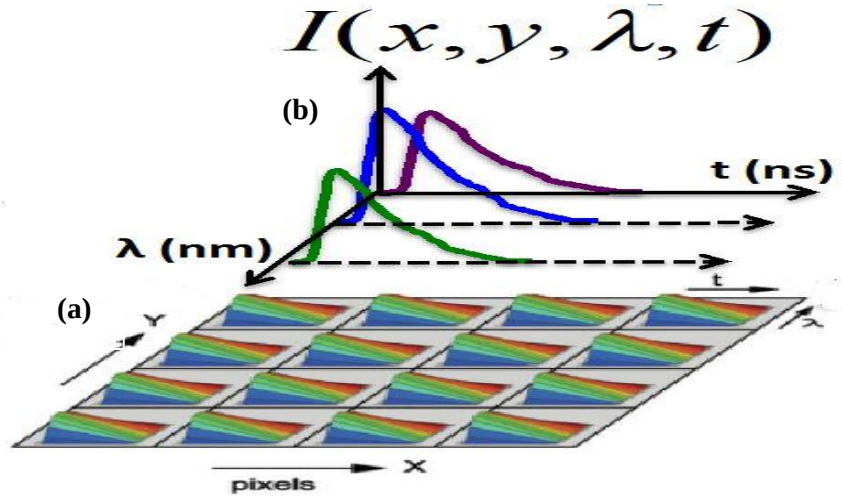

Figure 3 Composition of FLIM data. (a) Actual FLIM data; (b) Actual signals of a pixel.

In a FLIM image, the fluorescence intensity generated at each pixel (x,y) is measured as a function of different emission wavelengths or spectral bands (λ) and a function of time (t), as depicted in Figure 3. The spectral bands can be divided based on the targeted fluorophores. We chose the 390±20 nm, 452±22.5 nm, and >500 nm bands to distinguish emission from three tissue endogenous fluorophores: collagen, NADH, and FAD, respectively. The signals of the three channels in a pixel are visualized in Figure 3 (b). I(x, y, λ, t) indicates that each intensity is determined by the pixel's location (x, y), wavelength (λ), and time (t). Figure 4 is an enlarged figure that is similar to Figure 3 (b) with real measured data.

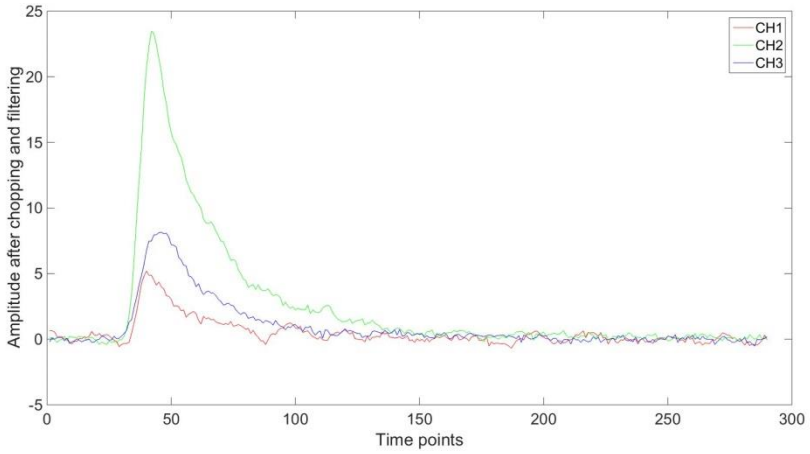

Figure 4 Enlarged actual signals of a pixel in FLIM data.

In the time-domain FLIM acquisition, the measured decay at a given pixel of the FLIM image represents the convolution of the instrument response with the fluorescence Impulse Response Function (IRF) of the sample at that pixel. Therefore, we need to implement a deconvolution procedure to estimate the real fluorescence decay. We use the previously reported Laguerre FLIM deconvolution algorithm [20] to estimate the intrinsic fluorescence decays at each pixel. Typically, the fluorescence decay *y(n)* is given by the convolution of the IRF *h(n)* with the instrument response *x(n)*: $y(n) = T \cdot \sum_{m=0}^{K-1} h(m)x(n-m),\ \ n = 0,1,\dots,N-1$, where T is the sampling interval, N is the number of measured samples, and K determines the time length of the IRF [21]. The Laguerre deconvolution technique uses a set of discrete Laguerre functions (DLF) as an orthonormal basis to represent IRF: $h(n) = T \cdot \sum_{j=0}^{L-1} c_j^{\alpha} b_j^{\alpha}(n)$, where c_j^{α} are the Laguerre Coefficients (LC), which are to be estimated from the input-output data, b_j^{α}

denotes the *j*-th order orthonormal DLF, and *L* is the number DLFs used to model the IRF[21]. The Laguerre parameter α determines the rate of asymptotic decline of the DLFs. The larger the α, the longer a DLF is spread over time. For a given pixel, the measured fluorescence decay *y(n)* can thus be written as: $y(n) = \sum_{j=0}^{L-1} c_j^{\alpha} v_j^{\alpha}(n)$, where: $v_j^{\alpha}(n) = \sum_{m=0}^{K-1} b_j^{\alpha}(m) x(n-m)$. After applying the Laguerre FLIM deconvolution algorithm, three deconvolved decays ($h_1(t), h_2(t), h_3(t)$) are acquired per pixel, each corresponding to the three emission bands described above. Figure 5 shows the deconvolved decays of a sample pixel. The decay profiles (h(t)) at each pixel are integrated to generate absolute fluorescence intensity maps corresponding to each wavelength band (I_1, I_2 and I_3). The normalized intensities (In) are calculated at each pixel as: $\mathrm{In1} = \frac{I_1}{I_1+I_2+I_3}$, $\mathrm{In2} = \frac{I_2}{I_1+I_2+I_3}$ and $\mathrm{In3} = \frac{I_3}{I_1+I_2+I_3}$. The average fluorescence lifetime (dlifetime) and $\frac{1}{e}$ lifetime (dtau) for each pixel are calculated using their general mathematical definitions [22]. Thus, the FLIM processing results in a set normalized intensity, average lifetime and 1/e lifetime maps, as shown in Figure 6 for a sample FLIM data set.

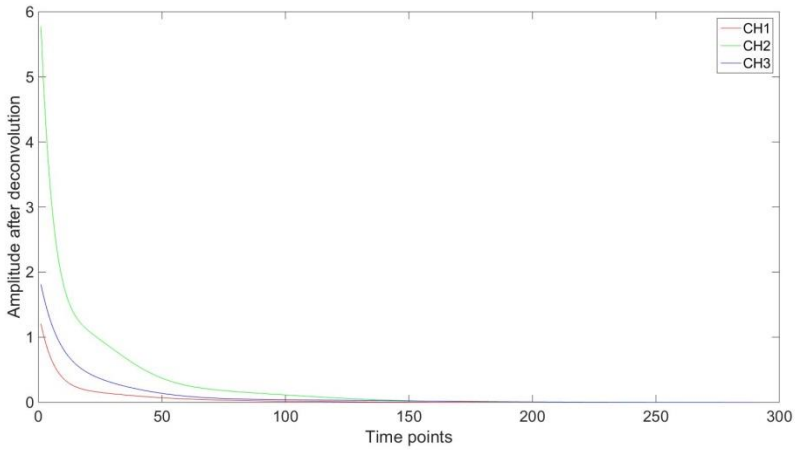

Figure 5 Deconvolved fluorescence decays.

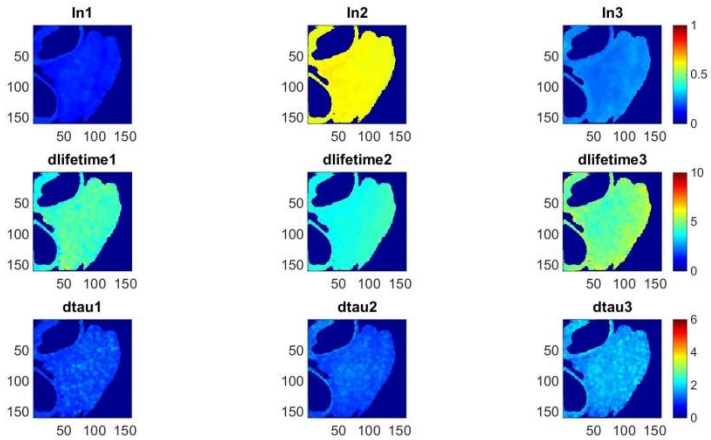

Figure 6 Final results of FLIM data processing. "In" means normalized intensity map, "dlifetime" means average lifetime map, and "dtau" means $\frac{1}{e}$ lifetime map. The number after In, dlifetime, and dtau indicates the channel number.

2.4 Feature pre-selection

The nine FLIM parameters estimated at each pixel of the image (as shown in Figure 6), represents FLIM features that can potentially be used for classifying the different types

of oral lesions, in particular: benign lesions (B), from pre-cancerous or dysplastic lesions (D) and squamous cell carcinomas (S). As a first approach, we used the pixel median value of each of these FLIM parameters as classification features. In order to identify those features useful for classification, we plotted the median FLIM parameter value of each lesion (Disease red markers) and its contralateral imaged side (Normal blue markers), as in Figure 7.

The data from the gingiva lesions are shown in Figure 7(a). It can be observed that the median value for In1 is virtually the same for the benign lesions and its corresponding contralateral normal area, while the cancerous lesions showed smaller median In1 values with respect to its corresponding contralateral normal area. It can also be observed that the median value for In3 is virtually the same for the benign lesions and its corresponding contralateral normal area, while the cancerous lesions showed larger median In3 values with respect to its corresponding contralateral normal area. Thus, two features were selected for classifying benign from cancerous gingival lesions: 1) difference in In1 between lesion and contralateral normal side, 2) difference in In3 between lesion and contralateral normal side.

The data from the tongue lesions are shown in Figure 7(b). It can be observed that the median value for In3 for the cancerous lesions was higher than for most benign and dysplastic lesions. It can also be observed that the median value for In3 is virtually the same for the benign lesions and its corresponding contralateral normal area, while most

dysplastic and cancerous lesions showed larger median In3 values with respect to its corresponding contralateral normal area. It can also be observed that the median value for dlifetime1, dlifetime2 and dtau2 is virtually the same for the benign lesions and its corresponding contralateral normal area, while most dysplastic and cancerous lesions showed smaller median values with respect to its corresponding contralateral normal area. Thus, five features were selected for classifying benign from cancerous tongue lesions: 1) In3, 2) difference in In3 between lesion and contralateral normal side, 3) difference in dlifetime1 between lesion and contralateral normal side, 4) difference in dlifetime2 between lesion and contralateral normal side, and 5) difference in dtau2 between lesion and contralateral normal side.

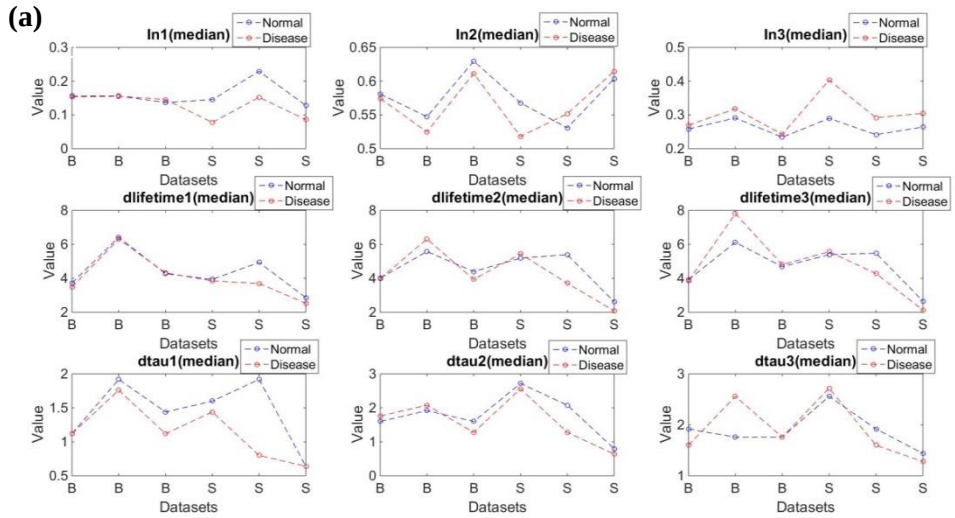

Figure 7 Median plots from gingiva and tongue. (a) Median plot of gingiva datasets. B indicates benign and S indicates SCC; (b) Median plot of tongue datasets. B indicates benign, D indicates dysplasia, and S means SCC.

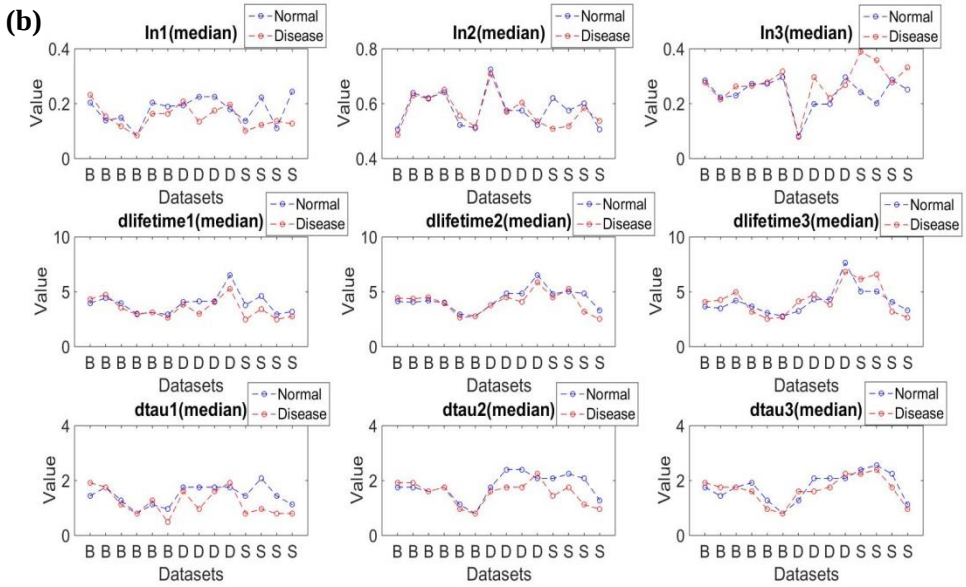

Figure 7 Continued.

In summary, based on these observations, a total of six FLIM parameters were pre-selected as potential discriminatory features for classifying benign from pre-cancerous and cancerous lesions.

2.5 Classification problems

The main objective of this work was to develop classification algorithms aiming to discriminate benign lesions from either pre-cancerous or cancerous lesions, based on FLIM derived features. For this purpose, we defined three classification problems: 1) benign vs. cancerous gingiva lesions; 2) benign vs. dysplastic/cancerous tongue lesions; and 3) benign vs. dysplastic/cancerous gingiva and tongue lesions.

2.6 Feature selection

Starting from the six pre-selected FLIM based features, exhaustive searching [23] was applied to find an optimal subset of features for each classification problem. Exhaustive searching is a feature selection method that explores the entire search space by testing all possible candidate solution. If we have n candidates and want to find the best set with size k, the number of subsets to be evaluated is $\binom{n}{k}$. This shows that the number of iterations can be huge even for modest n and k. In our case, however, there were only six candidates to be considered; thus, the cost of exhaustive searching is low.

2.7 Classification algorithms

For evaluating each feature subset, we need classification rules to estimate the error. In this study, three different classification rules were considered: Linear Discriminant Analysis (LDA), 3-Nearest-Neighbor (3NN), and Support Vector Machines (SVM). They are described below.

2.7.1 Linear Discriminant Analysis (LDA) [23]

LDA is a method for finding a linear combination of features that separates data into a number of classes (two classes for our classification problems). The designed classifier is derived as

$$\varphi_n(x) = \begin{cases} 1, & {a_n}^T x + b_n > 0 \\ 0, & otherwise \end{cases}$$

where $a_n = \hat{\sum}^{-1} \cdot (u_1 - u_0)$, $b_n = -\frac{1}{2}(u_1 - u_0)^T \cdot \hat{\sum}^{-1} \cdot (u_1 + u_0)$ and x is the training feature vectors. In detail, $\hat{\sum}$ denotes the single covariance matrix, which is estimated by the pooled sample covariance matrix and u_1, u_0 are the mean feature vectors in each group.

2.7.2 3-Nearest-Neighbor (3NN) [23]

3NN is a nonparametric method that estimates class-conditional densities without making any distributional densities. This algorithm uses the majority vote of its neighbors. In other words, the class of an object is assigned to the most common class among its K-Nearest-Neighbor (KNN). It is a great way to assign weight to the contributions of the neighbors because nearer neighbors contribute more than distant ones. The parameter *k* must be changed according to the data. Larger *k* can reduce the effect of noise on the classification [24], while it produces less distinct boundaries between classes. In binary classification, it is helpful to choose k to be an odd number because of the case of tied votes. We considered k=1, 3, 5 and finally, 3-Nearest-Neighbor shows the best result because of its good speed and accuracy. The designed classifier is derived as

$$\varphi_n(x) = \begin{cases} 1, & \sum_{i=1}^{3} I_{\{Y_i=1\}} > \sum_{i=1}^{3} I_{\{Y_i=0\}} \\ 0, & otherwise \end{cases}$$

Here, x is the training feature vectors and y is the labels of that training feature vectors. $\varphi_n(x)$ is the designed classifier and $I_{\{Y_i=1\}}$ or $I_{\{Y_i=0\}}$ is the i-th nearest data point to x

that has the label 1 or label 0. Thus, the designed classifier will be 0 if the nearest neighbors of x have more 0 than 1 labels.

2.7.3 Support Vector Machines (SVM) [23]

SVM is a supervised learning model that analyzes data and recognizes patterns that is used for classification. This model separates two categories with a clear gap by making the gap as wide as possible. After that, data is mapped into that same space and predicted to belong to a category based on which side of the gap they fall into. The main reason why we employ this method is that SVM can perform non-linear classification using the kernel approach. In our study, we use linear and Gaussian kernels. The designed classifier by SVM with linear kernel is given by:

$$\varphi_n(x) = \begin{cases} 1, & if \sum_{i \in S} \lambda_i^* y_i x_i^T x - \frac{1}{n_s} \sum_{i \in S} \sum_{j \in S} \lambda_i^* y_i x_i^T x_j + \frac{1}{n_s} \sum_{i \in S} y_i > 0 \\ 0, & otherwise \end{cases}$$

Here, $\varphi_n(x)$ is the designed classifier and λ_i^* is the optimal Lagrange multiplier, which can be achieved from maximizing the dual Lagrange functional [23]. The set $\{(x_i, y_i); i \in S\}$ is the set of support vector points, x is the training feature vectors, y is the labels of that training feature vectors (±1, instead of 0 and 1), and n_s is the number of support vectors. The designed classifier by SVM with Gaussian kernel is given by:

$$\varphi_n(x) = \begin{cases} 1, & if \sum_{i \in S} \lambda_i^* y_i k(x_i, x) - \frac{1}{n_s} \sum_{i \in S} \sum_{j \in S} \lambda_i^* y_i k(x_i, x_j) + \frac{1}{n_s} \sum_{i \in S} y_i > 0 \\ 0, & otherwise \end{cases}$$

where $k(x_i, x_j)$ is a kernel function, and we only employ the Gaussian kernel, which is $k(x_i, x_j) = \exp(\frac{-|x_i - x_j|^2}{\sigma^2})$.

2.8 Performance quantification

To estimate the error at the feature selection stage, the Leave-One-Out Cross Validation (LOO) method [23] was used. LOO is a model evaluation method for measuring how results of the classification analysis are generalized to an independent dataset. To achieve this, one of the datasets is removed before training. After training, the removed dataset is used to test the performance of the learned classifier. The main goal of LOO is to achieve unbiased estimation results and also, explain the model generalization in an independent dataset.

For each classification problem, the training data consisted on the pixel median values of the FLIM features. Thus, for each lesion, we only have one training data. For LOO validation, we followed two approaches. For the first approach, denoted 'sample classification', the leave-out lesion was classified as a single sample based on the pixel median FLIM feature values. For the second approach, denoted 'pixel classification', each pixel of the leave-out lesion was classified based on each pixel FLIM feature values. The exhaustive feature search was performed based on the sample classification accuracy. In addition, the pixel classification performance was quantified in the LOO overall classification accuracy, and sensitive and specificity for classifying dysplastic or cancerous pixels.

3. RESULT

3.1 Histopathology evaluation

A total of 20 lesions were imaged, six from the gingiva, and the other 14 from the tongue. Table 1 shows a summary of histopathology distribution. To be specific, dysplasia means mild epithelial dysplasia and SCC means squamous cell carcinoma. All other pathological diagnoses are benign.

Location	Pathological Diagnosis
Tongue	Hyperparakeratosis and epithelial atypia with focal lichenoid features
Tongue	Fibroma
Tongue	Lichen planus
Tongue	Hyperparakeratosis, acanthosis,chronicmucositis and cellular atypia
Tongue	Non-specific ulcer
Tongue	Chronic mucositis with lichenoid features
Gingiva	Slight parakeratosis
Gingiva	Chronic mucositis with lichenoid features
Gingiva	Hyperparakeratosis and chronic mucositis
Tongue	Mild epithelial dysplasia
Tongue	Mild epithelial dysplasia
Tongue	Mild epithelial dysplasia
Tongue	Mild epithelial dysplasia
Tongue	Squamous cell carcinoma
Tongue	Squamous cell carcinoma
Tongue	Squamous cell carcinoma
Tongue	Squamous cell carcinoma
Gingiva	Squamous cell carcinoma
Gingiva	Squamous cell carcinoma
Gingiva	Squamous cell carcinoma

Table 1 Summary of histopathology distribution.

3.2 Classification – Gingiva

A total of six gingiva lesions were imaged, from which three were benign and three were cancerous. Starting from the six FLIM based featured pre-selected before, the exhaustive searching was applied to find the best feature set among all the combinations that would classify the benign from the cancerous lesions. Based on the exhaustive searching, an LDA classification rule with the following two features was identified as optimal: 1) difference in In1 between lesion and contralateral normal side, 2) difference in In3 between lesion and contralateral normal side.

Figure 8 is a scatter plot with the best features and classification rule for the gingiva classification. It shows that all datasets were correctly classified. Table 2 shows the LOO sample-level confusion matrix of the gingiva area, indicating that all samples were classified correctly.

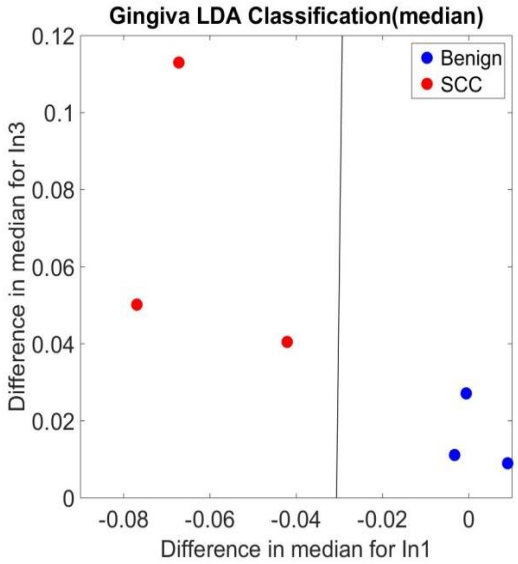

Figure 8 Scatter plot based on gingiva training datasets. Blue dots are benign, and red dots are SCC. The X-axis is the first feature, and the Y-axis is the second feature used in the gingiva. The black line is a decision boundary from LDA that distinguishes benign from SCC. In this figure, LDA was trained with all samples for better visualization, even if we actually apply LOO at the feature selection stage.

LDA(sample level)

	Benign (Green)	**SCC (Red)**	**Total Samples**
Benign	3 (100%)	0 (0%)	3
SCC	0 (0%)	3 (100%)	3

Table 2 LOO sample-level confusion matrix of gingiva area.

Next, we performed a pixel classification with the same features and classification rule as the sample classification. To achieve this, pixel median values of the FLIM features were used for training and each pixel value of FLIM feature from the leaved-out FLIM image was used for testing the trained classification rule. Figure 9 shows the classification maps derived from two selected features using the LDA classification rule. Table 3 shows the LOO pixel-level confusion matrix of the gingiva, indicating high sensitivity and very high specificity for cancer detection.

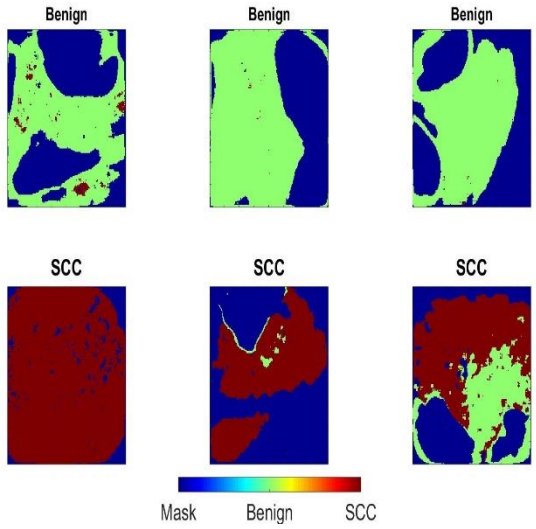

Figure 9 Classification maps of gingiva. The green color means benign, the red color means SCC, and the blue color (mask) represents the portion excluded from the classification process. Benign should be mostly green, while SCC is mainly red.

LDA(pixel level)

	Benign (Green)	SCC (Red)	Total Pixels
Benign	43,660 (98.7%)	582 (1.3%)	44,242
SCC	6,947 (13.3%)	45,098 (86.7%)	52,045

Table 3 LOO pixel-level confusion matrix of gingiva area. It shows that sensitivity for SCC is 86.7%, while specificity for SCC is 98.7%.

3.3 Classification – Tongue

In tongue, 14 lesions were imaged, from which six were benign, four were precancerous, and four were cancerous. As in the gingiva, the exhaustive searching was used to find the best feature set among all the combinations of candidates that would distinguish benign from the precancerous and cancerous lesions. From the exhaustive searching, SVM Gaussian kernel with the following three features was determined as optimal: 1) difference in In3 between lesion and contralateral normal side, 2) In3, and 3) difference in dlifetime2 between lesion and contralateral normal side.

Figure 10 is a scatter plot with the best features and classification rule of the tongue classification. It shows that all datasets were correctly distinguished. Table 4 shows the LOO sample-level confusion matrix of the tongue area, representing that all samples are classified perfectly.

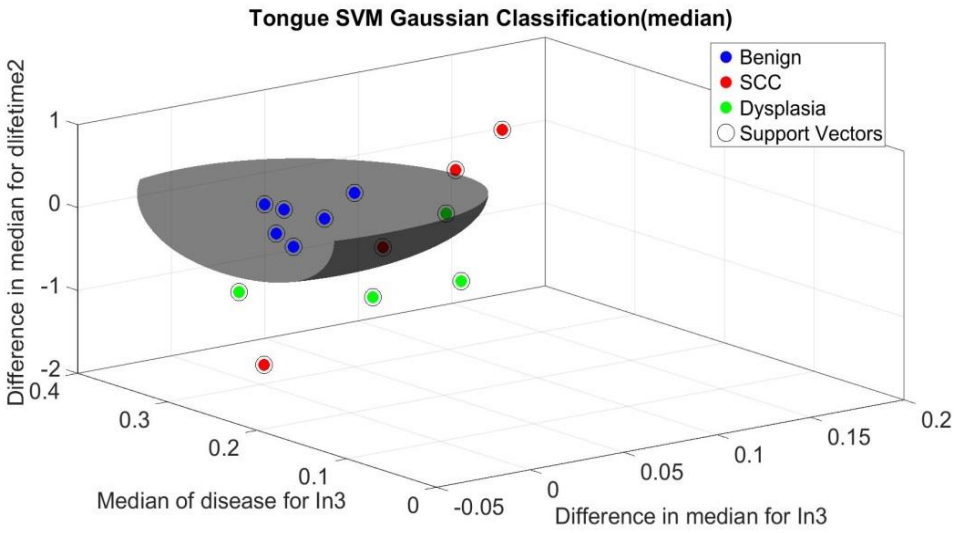

Figure 10 Scatter plot based on tongue training datasets. Blue dots are benign, green dots are dysplasia, and red dots are SCC. Additionally, the black circles that surround the dots are support vectors. The X-axis is the first feature, the Y-axis is the second feature, and the Z-axis is the third feature used in the tongue area. The black hyperplane is a decision boundary from SVM Gaussian kernel that distinguishes benign from dysplasia and SCC. In this figure, the classification rule was trained with all samples for better visualization, even if we actually apply LOO at the feature selection stage.

SVM_Gaussian(sample level)

	Benign (Green)	Dysplasia/SCC (Red)	Total Samples
Benign	6 (100%)	0 (0%)	6
Dysplasia	0 (0%)	4 (100%)	4
SCC	0 (0%)	4 (100%)	4

Table 4 LOO sample-level confusion matrix of tongue area.

Next, we conducted a pixel classification with the same features and classification rule that were the best in the sample classification. As before, pixel median values of the FLIM features were applied for training and each pixel value of FLIM feature from the leaved-out FLIM image was applied for testing the trained classification rule. Figure 11 shows the classification maps derived from three selected features using the SVM Gaussian kernel. Table 5 represents the LOO pixel-level confusion matrix of the tongue datasets, indicating high sensitivity and specificity for precancer and cancer detection.

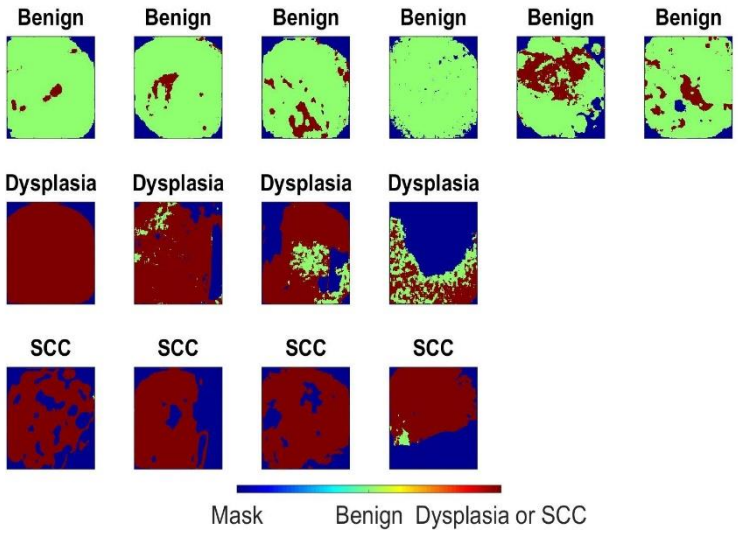

Figure 11 Classification maps of tongue. The green color means benign, the red color means Dysplasia/SCC, and the blue color (mask) shows the portion excluded from the classification procedure. Benign should be mainly green, while Dysplasia/SCC must be mostly red.

SVM_Gaussian(pixel level)

	Benign (Green)	Dysplasia/SCC (Red)	Total Pixels
Benign	123,845 (90%)	13,825 (10%)	137,670
Dysplasia/ SCC	9,762 (6.8%)	134,692 (93.2%)	144,454

Table 5 LOO pixel-level confusion matrix of tongue area. It shows that sensitivity for Dysplasia/SCC is 93.2%, while specificity for Dysplasia/SCC is 90%.

3.4 Classification – Gingiva and Tongue

In the entire area, a total of 20 lesions were collected, from which nine were benign, four were precancerous, and seven were cancerous. Beginning from the six FLIM based featured pre-selected before, the exhaustive searching was applied to discover the best feature combination among all the combinations that would be helpful to classify the benign from the precancerous and cancerous lesions. Based on the exhaustive searching, SVM Gaussian kernel with the following three features were identified as optimal: 1) difference in In3 between lesion and contralateral normal side, 2) In3, and 3) difference in dlifetime2 between lesion and contralateral normal side.

Figure 12 is a scatter plot with the best features and classification rule for the entire classification. It indicates that all datasets were well classified, except for one dysplasia.

Table 6 represents the LOO sample-level confusion matrix of the entire area, indicating that all samples were classified correctly, except for one dysplasia.

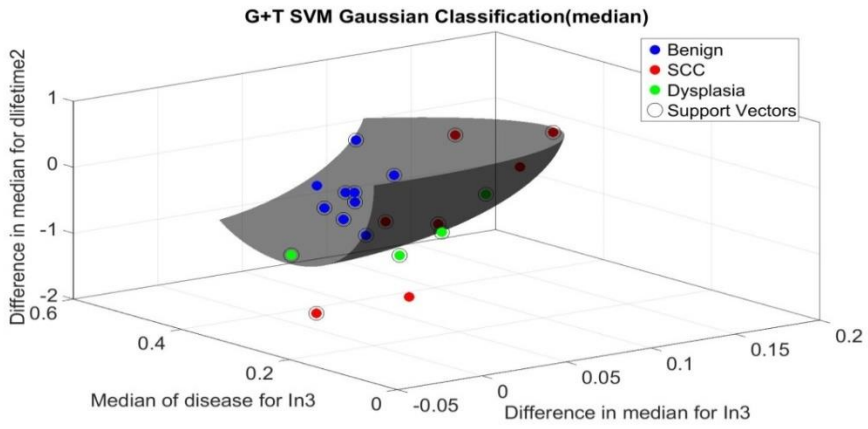

Figure 12 Scatter plot based on total training datasets. Blue dots are benign, green dots are dysplasia, and red dots are SCC. In addition, the black circles that surround the dots are support vectors. The black hyperplane is a decision boundary from SVM Gaussian kernel that separates benign from dysplasia and SCC. In this figure, SVM Gaussian kernel was trained with all samples for better visualization, even if we actually apply LOO at the feature selection stage.

SVM_Gaussian(sample level)

	Benign (Green)	**Dysplasia/SCC (Red)**	**Total Samples**
Benign	9 (100%)	0 (0%)	9
Dysplasia	1 (25%)	3 (75%)	4
SCC	0 (0%)	7 (100%)	7

Table 6 LOO sample-level confusion matrix of entire area.

As before, we performed a pixel classification with the same features and classification rule as the sample classification. To achieve this, pixel median values were used for training and each pixel value from the leaved-out FLIM image was applied for testing the trained classification rule. Figure 13 shows the classification maps resulted from three selected features using the SVM Gaussian kernel. Table 7 represents the LOO pixel-level confusion matrix of the entire datasets, showing high sensitivity and very high specificity for precancer and cancer detection.

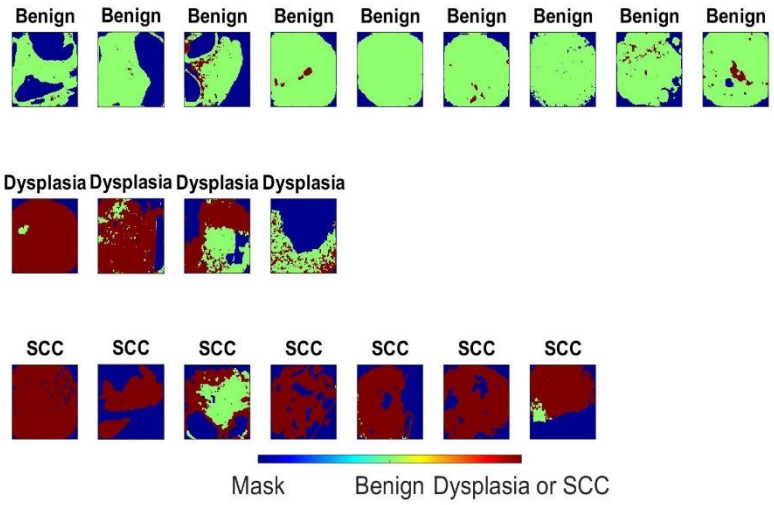

Figure 13 Classification maps of gingiva and tongue. The green color means benign, the red color means Dysplasia/SCC, and the blue color (mask) shows the portion excluded from the classification process. Benign should be mostly green, while Dysplasia/SCC is mainly red.

SVM_Gaussian(pixel level)

	Benign (Green)	Dysplasia/SCC (Red)	Total Pixels
Benign	176,650 (97.1%)	5,262 (2.9%)	181,912
Dysplasia/ SCC	28,076 (14.3%)	168,423 (85.7%)	196,499

Table 7 LOO pixel-level confusion matrix of entire area. It shows sensitivity for Dysplasia/SCC is 85.7%, while specificity for Dysplasia/SCC is 97.1%.

4. CONCLUSIONS

To the best of our knowledge, this is the first demonstration of endogenous FLIM endoscopic imaging for automated detection of a wide range of oral benign lesions (including inflammation) from oral pre-cancerous (mild-dysplasia) and cancerous (SCC) lesions. Our results indicate that both fluorescence intensity and fluorescence lifetime derived features are important in order to achieve decent lesion classification performance. Fluorescence lifetime was in particular relevant to distinguish pre-cancerous lesions from benign lesions.

We should recognize, however, that our current database is limited to a few lesions in the gingiva and tongue areas of the oral cavity, including the vestibule, palate and mucosa. However, our preliminary results are encouraging and our ongoing efforts are focused on further validating on a larger database this promising technology for early detection of oral epithelial cancer.

REFERENCES

[1] Ramanujam, N., *Fluorescence spectroscopy of neoplastic and non-neoplastic tissues.* Neoplasia, 2000. 2(1-2): p. 89-117.

[2] Georgakoudi, I., Jacobson, B.C, Muller, M.G, Sheets, E.E, Badizadegan,K, et al., *NAD(P)H and collagen as* in vivo *quantitative fluorescent biomarkers of epithelial precancerous changes.* Cancer Research, 2002. 62(3): p. 682-7.

[3] Wu, Y.C. and J.N.Y. Qu, *Autofluorescence spectroscopy of epithelial tissues.* Journal of Biomedical Optics, 2006. 11(5): 054023.

[4] Pavlova, I., Williams, M., El-Naggar, A., Richards-Kortum, R., Gillenwater, A., *Understanding the biological basis of autofluorescence imaging for oral cancer detection: high-resolution fluorescence microscopy in viable tissue.* Clinical Cancer Research, 2008. 14(8): p. 2396-2404.

[5] Harris, D.M. and J. Werkhaven, *Endogenous porphyrin fluorescence in tumors.* Lasers in Surgery and Medicine, 1987. 7(6): p. 467-472.

[6] Betz, C.S., Mehlmann, M., Rick, K., Stepp, H., Grevers, G, et al., *Autofluorescence imaging and spectroscopy of normal and malignant mucosa in patients with head and neck cancer.* Lasers in Surgery and Medicine, 1999. 25(4): p. 323-334.

[7] Roblyer, D., Kurachi, C., Stepanek, V., Williams, M.D., El-Naggar, A.K., et al., *Objective detection and delineation of oral neoplasia using autofluorescence imaging.* Cancer Prevention Research, 2009. 2(5): p. 423-431.

[8] Bedard, N., Schwarz, R.A., Hu, A., Bhattar, V., Howe, J., et al., *Multimodal snapshot spectral imaging for oral cancer diagnostics: a pilot study.* Biomedical Optics Express, 2013. 4(6): p. 938-949.

[9] Lee, J.S.,Shuhatovich, O., Price, R., Pikkula, B., Follen, M., et al., *Design and preliminary analysis of a study to assess intra-device and inter-device variability of fluorescence spectroscopy instruments for detecting cervical neoplasia.* Gynecologic Oncology, 2005. 99(3 Suppl 1): p. S98-111.

[10] Marin, N.M., MacKinnon, N., MacAulay, C., Chang, S.K., Atkinson, E.N., et al., *Calibration standards for multicenter clinical trials of fluorescence spectroscopy for* in vivo *diagnosis.* Journal of Biomedical Optics, 2006. 11(1): 014010.

[11] Pikkula, B.M., Shuhatovich, O., Price, R.L., Serachitopol, D.M., Follen, M., et al., *Instrumentation as a source of variability in the application of fluorescence spectroscopic devices for detecting cervical neoplasia.* Journal of Biomedical Optics, 2007. 12(3): 034014.

[12] Morgan, C.G. and A.C. Mitchell, *Fluorescence lifetime imaging: an emerging technique in fluorescence microscopy.* Chromosome Res, 1996. 4(4): p. 261-263.

[13] Lakowicz, J.R., *Principles of fluorescence spectroscopy*. 2nd ed. 1999, New York: Kluwer Academic/Plenum Publishers. p. 95-140.

[14] Ramanujan, V.K., Zhang, J.H., Biener, E., Herman, B., *Multiphoton fluorescence lifetime contrast in deep tissue imaging: prospects in redox imaging and disease diagnosis*. Journal of Biomedical Optics, 2005. 10(5): 051407.

[15] Chen, H.M., Chiang, C.P., You, C., Hsiao, T.C., Wang, C.Y., et al., *Time-resolved*

autofluorescence spectroscopy for classifying normal and premalignant oral tissues. Lasers in Surgery and Medicine, 2005. 37(1): p. 37-45.

[16] Skala, M.C., Squirrell, J.M., Vrotsos, K.M., Eickhoff, J.C., Gendron-Fitzpatrick, A., et al., *Multiphoton microscopy of endogenous fluorescence differentiates normal, precancerous, and cancerous squamous epithelial tissues.* Cancer Research, 2005. 65(4): p. 1180-6.

[17] Skala, M.C., Riching, K.M., Gendron-Fitzpatrick, A., Eickhoff, J., Eliceiri, K.W., et al., In vivo *multiphoton microscopy of NADH and FAD redox states, fluorescence lifetimes, and cellular morphology in precancerous epithelia.* Proceedings of the National Academy of Sciences, 2007. 104(49): p. 19494-19499.

[18] Sun, Y., Phipps, J.E., Meier, J., Hatami, N., Poirier, B., et al., *Endoscopic fluorescence lifetime imaging for* in vivo *intraoperative diagnosis of oral carcinoma.* Microscopy and Microanalysis, 2013. FirstView: p. 1-8.

[19] Cheng, S., Cuenca, R.M., Liu, B., Malik B.H., Jabbour, J.M., et al., *Handheld multispectral fluorescence lifetime imaging system for* in vivo *applications.* Biomedical Optics Express, 2014. 5(3): p. 921-931.

[20] Jo, J.A., Marcu, L., Fang, Q., Papaioannou, T., Qiao, J.H., et al., *New methods for time-resolved fluorescence spectroscopy data analysis based on the laguerre expansion technique--applications in tissue diagnosis.* Methods of Information in Medicine, 2007. 46(2): p. 206-11.

[21] Jo, J.A., Fang, Q., and Marcu, L., *Ultrafast method for the analysis of*

fluorescence lifetime imaging microscopy data based on the laguerre expansion technique. IEEE Journal of Quantum Electronics, 2005. 11(4): p. 835-845.

[22] Lakowicz, J.R., *Principles of fluorescence spectroscopy.* Chemistry in Britain, 1984. 20(5): p. 442-442.

[23] Duda, R.O., Hart, P.E., and Stork, D.G., *Pattern classification*. 2nd ed. 2001, New York: Wiley. xx, p. 654.

[24] Everitt, B.S., Landau, S., Leese, M., and Stahl, D., *Cluster analysis*. 5th ed. 2011, Chichester, West Sussex, U.K.: Wiley. xii, p. 330.