Chapter 4

Clinical trial for non-invasive detection of oral lesions by point monitoring of laser-induced autofluorescence

Clinical trial for non-invasive detection of oral lesions by point monitoring of laser-induced autofluorescence

4.1 Introduction

The potential of utility of laser-induced autofluorescence (LIAF) technique for diagnosis of oral lesions in a clinical situation is presented. Towards this, we have recorded the *in vivo* laser-induced autofluorescence spectra of oral lesions from 61 patients and 30 healthy volunteers in the 420-720 nm wavelength regions with diode laser excitation. Linear discriminant analysis (LDA) based on leave-one-out (LOO) method of cross validation was performed on spectral data for tissue characterization and the results are presented. The discriminant function scores determined by LDA-LOO method were used to find a cut-off value for discriminating normal from hyperplasia, hyperplasia from dysplasia and dysplasia from squamous cell carcinoma (SCC) lesions and the sensitivity and specificity for different lesion pairs were determined from the scatter plot of discriminant function scores and the results are detailed.

4.2. Materials and Methods

4.2.1 Study Population and Protocol

Autofluorescence spectra were recorded from 30 healthy volunteers with no clinically observable lesions and 61 patients having different grades of malignancy. An experienced physician specialized in radiation oncology identifies lesions for spectral studies and records its visual imprint. All the patients included in this study had prolonged habits of either pan

chewing, smoking or alcohol consumption, while healthy subjects maintained good oral hygiene with none of the above habits. Before measurements, all patients and healthy volunteers were asked to rinse their mouth with 0.9% saline solution for two minutes in order to reduce the effect of recently consumed food. They were also provided with protective goggles to shield their eyes against laser light. After 15 sets of spectral measurements from a typical site covering an area of 6 mm dia., a punch biopsy (2x2 mm approximately in size) was taken from the central portion of the measurement site. The biopsy slides were prepared and classified by an experienced pathologist who was blinded to the autofluorescence spectral results. In the case of healthy volunteers, visual impression was carried out instead of biopsy. After measurements, the spectroscopic results were correlated with histopathological findings.

4.2.2 Data processing & Statistical Analysis

Autofluorescence spectral data recorded from different oral cavity sites of patients/volunteers categorized based on pathological reports as normal (30 healthy volunteers), hyperplasia (19 patients), dysplasia (16 patients), and SCC (26 patients) were included in this study. 59 samples randomly selected and belonging to these categories with known group membership (pathological results) were used as training/prediction set and the remaining 32 samples with unknown group membership were used for validation. Data preprocessing methods such as normalization, mean-scaling and normalization cum mean-scaling were used to find out whether there is any spectral enhancement due to preprocessing (Ramanujam et al., 1996a). Spectra from each patient was averaged and then normalized individually to the maximum fluorescence intensity of each spectrum. Mean-scaling was performed by calculating the mean of each sample and then subtracting this mean from each spectrum, and the resultant value represents the difference in fluorescence at a particular site with respect to the average spectrum.

Since the fluorescence intensity data set extends from 420- 720 nm (with 3 values for each nm), we used PCA to identify the orthogonal components of the spectra with maximum variance within the complete data set. LDA was then performed on the extracted significant principal components to determine the discriminant function scores. Each principal component is co-related to the spectral variables of the original fluorescence emission spectra and provides insight into the spectral features that contribute to the classification (Ramanujam et al., 1996a, Ramanujam et al., 1996b). Independent student's t-test was carried out for testing the significance of mean PC scores between different tissue categories. The extracted significant principal components with p-value <0.005 were used as input variables for discriminant analysis, which produces a classification table to assess how well the discriminant function works. In order to assess the significance of the discriminant function

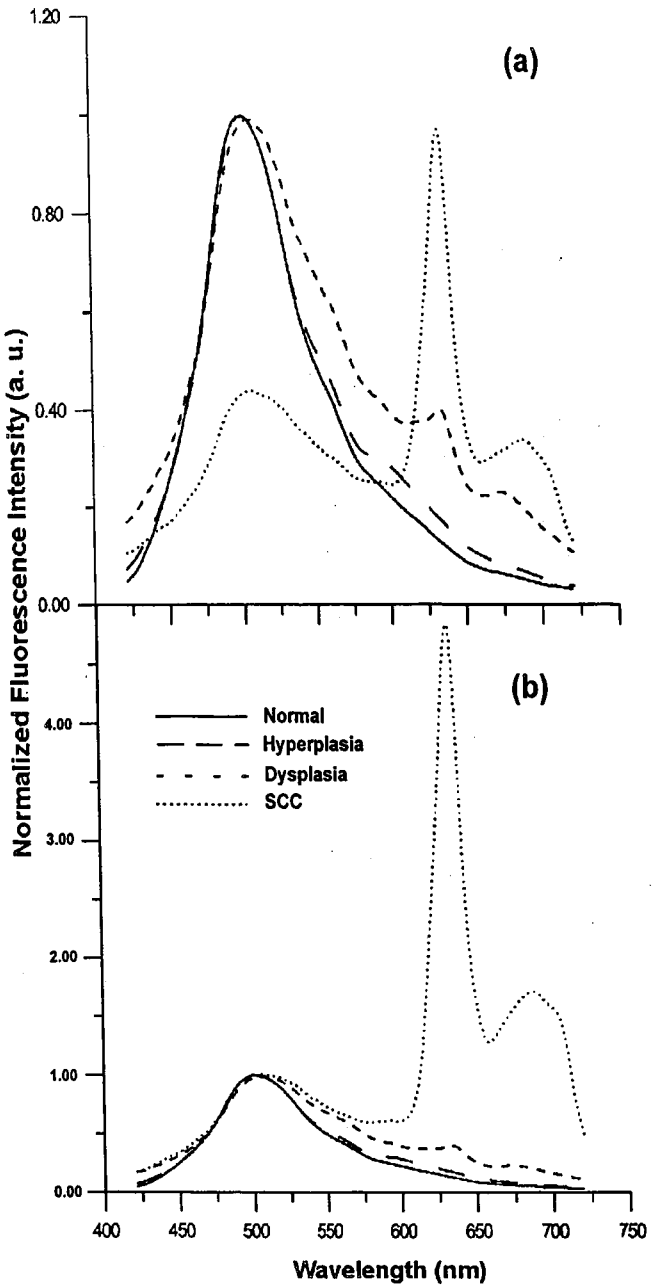

Fig. 4.1. LIAF emission from the oral mucosa of 61 patients and 30 healthy volunteers. (a) Normalized to maximum emission intensity of different lesions (b) Normalized to the intensity at 500 nm. Normal spectra show the average of 15 measurements carried out at 11 anatomical sites in 30 healthy volunteers, whereas hyperplasia, dysplasia and SCC spectra are the mean of 15 measurements each in 19, 16 and 26 patients, respectively.

scores, the Wilks Lambda test was used. Since the discriminant scores are predictors of group membership for classifying observations, we have performed a blind-test using the remaining 32 samples to assess the suitability of the training data set.

4.3. RESULTS

4.3.1 Spectral features

In vivo autofluorescence spectral measurements were carried out as per the protocol approved by the Ethics Committee of the institution on consenting healthy volunteers and patients. Figure 4.1 shows the mean fluorescence spectra of 15 measurements on each site for different tissue categorizes normalized at their peak emission intensity (Fig 4.1.a) and the spectra normalized for autofluorescence intensity at 500 nm (Fig 4.1.b). Marked differences in spectral features are observed between healthy and diseased tissues. The broad autofluorescence peak at 500 nm arising from endogenous fluorophore is characteristic of all epithelial tissues. The 635 nm peak is more prominent in lesions diagnosed as SCC as compared to dysplastic and hyperplastic tissues. In addition two peaks are also seen at 685 and 705 nm that are absent in healthy tissues. Another notable feature is the broadening of the 500 nm peak towards red wavelength region in dysplastic tissues.

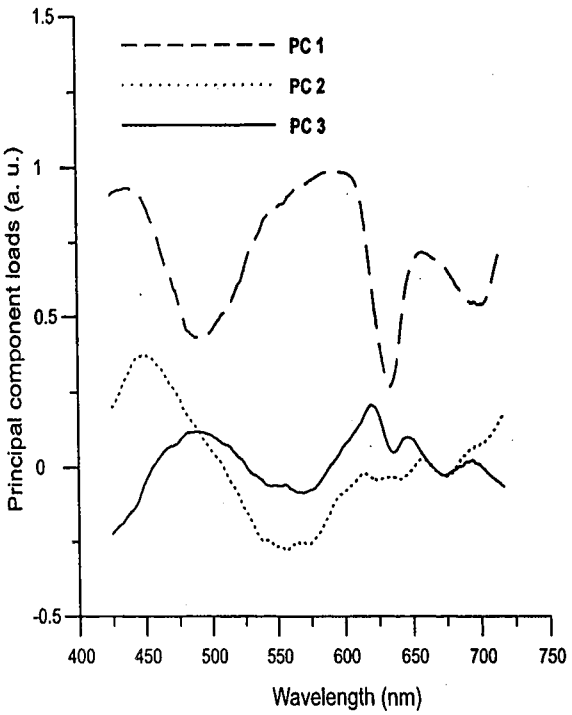

Fig. 4.2. First three principal components of different lesion groups included in the analysis. Each averaged spectrum was normalized with respect to the emission peak intensity.

The component-loading plot (Fig 4. 2) shows loadings of the first three principal components (PCs) for all the lesions studied and depicts the correlation between each

principal component and variables of the normalized fluorescence spectrum. PC1 represents the mean autofluorescence spectrum for all lesions. PC2 and PC3 look similar to the porphyrin peaks at 635 and 705 nm and also show dips at 545 and 575 nm due to oxygenated hemoglobin absorption. These three PCs jointly give a variance of 98.6% to the spectra.

4.3.2. Lesion classification

Pair-wise discriminant analysis was performed on the extracted significant principal components from the training set. Scatter plots were drawn based on the discriminant function scores obtained (Fig. 4.3a-c). Since the scores obtained by LDA can either be negative or positive, the scatter plot for each pair is given separately. The cut-off value in the scatter plot, which is the weighted mean of the paired values, is used to classify different lesions. The cut-off values for normal-hyperplastic, hyperplastic-dysplastic, SCC-dysplastic pairs are -0.156, 0.393 and -0.340 respectively. Diagnostic accuracies such as sensitivity, specificity, positive and negative predictive values for each pair were calculated by correlating the position of discriminant

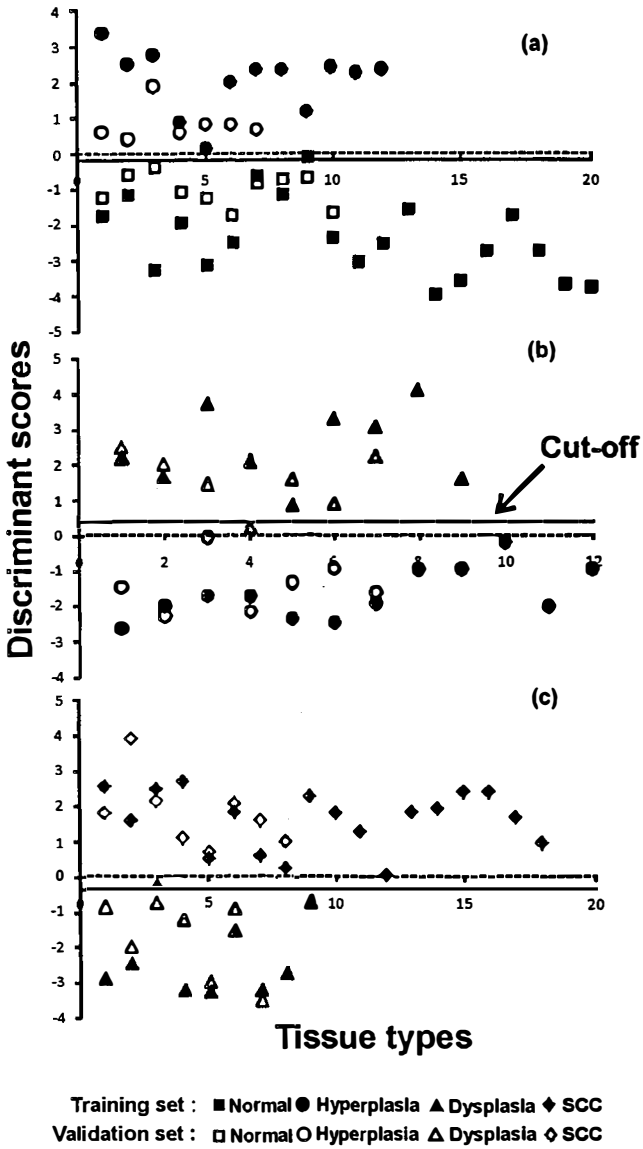

Fig. 4.3. Pair-wise scatter plot based on discriminant function scores for 30 healthy volunteers and 61 patients with different lesion types. (a) normal-hyperplasia, (b) hyperplasia-dysplasia, and (c) dysplasia-SCC. The solid symbols represent the results of training data set and the open symbols relate to the validation data set.

function scores for each lesion in the scatter plot with the corresponding histopathological result. For the prediction (training/standard) data set 100% sensitivity was obtained for normal-hyperplasia, hyperplasia-dysplasia and dysplasia-SCC pairs and the corresponding specificities were 95%, 100% and 89%, respectively (Table 4. 1). In order to test the reliability of the classification procedure used, a blind-test was carried out in 32 patients with unknown group membership. Discriminant function scores of the blind test data were inserted into the scatter plot of the training set for validation and the results were correlated with histopathological findings. It was observed that the developed algorithm could classify seventeen samples of normal-hyperplastic pairs, thirteen samples out of fourteen dysplastic-hyperplastic pairs (with one dysplastic case misclassified as hyperplasia) and fifteen samples dysplastic-SCC pairs correctly. This leads to a specificity of 100% and sensitivity of 100%, 86%, and 100% respectively for normal-hyperplasia, hyperplasia-dysplasia and dysplasia-SCC lesion pairs in the blind-test (Fig. 4.3a-c).

Table 4.1 Overall diagnostic accuracies obtained for different lesion pairs consisting of 59 samples in the training (prediction) set and 32 samples in the validation (blind test) data set

Lesion Pairs	Normal vs. Hyperplasia					Hyperplasia vs. Dysplasia					Dysplasia vs. SCC				
	Se (%)	Sp (%)	Acc (%)	PPV (%)	NPV (%)	Se (%)	Sp (%)	Acc (%)	PPV (%)	NPV (%)	Se (%)	Sp (%)	Acc (%)	PPV (%)	NPV (%)
Training set	100	95	96	92	100	100	100	100	100	100	100	89	96	95	100
Validation set	100	100	100	100	100	86	100	93	100	88	100	100	100	100	100
Overall	100	98	98	96	100	98	100	97	100	94	100	95	98	98	100

Se: Sensitivity; Sp: Specificity; Acc: Accuracy ((true positive + true negative)/(positive + negative)); PPV: Positive Predictive value; NPV: Negative Predictive Value

The overall sensitivity, specificity, accuracy, positive predictive value (PPV) and negative predictive value (NPV) calculated from the scatter plots (Figure 4. 3a-c) are given in Table 4. 1. We were able to discriminate premalignant dysplastic lesions from hyperplasia with 98% sensitivity, and 100% specificity, whereas for discriminating hyperplasia from normal and SCC from dysplasia, the sensitivities were 100% with corresponding specificities of 98% and 95%.

4.4. Discussion

The autofluorescence peak at 500 nm has been reported as due to emission from endogenous fluorophores like NADH, FAD, collagen, elastin and amino acids, while the emissions at 635 and 705 nm are from enhanced occurrence of PpIX in malignant tissues (Mallia et al., 2008a; Richards-Kortum et al., 1996; Inaguma et al., 1999). The additional peak seen at 685 nm in SCC and dysplastic lesions (Fig. 4.1) is due to the accumulation of corporporphyrin III, a constituent of the heme synthesis pathway in malignant tissues (Mallia et al., 2008a, Moesta et al.,2001).

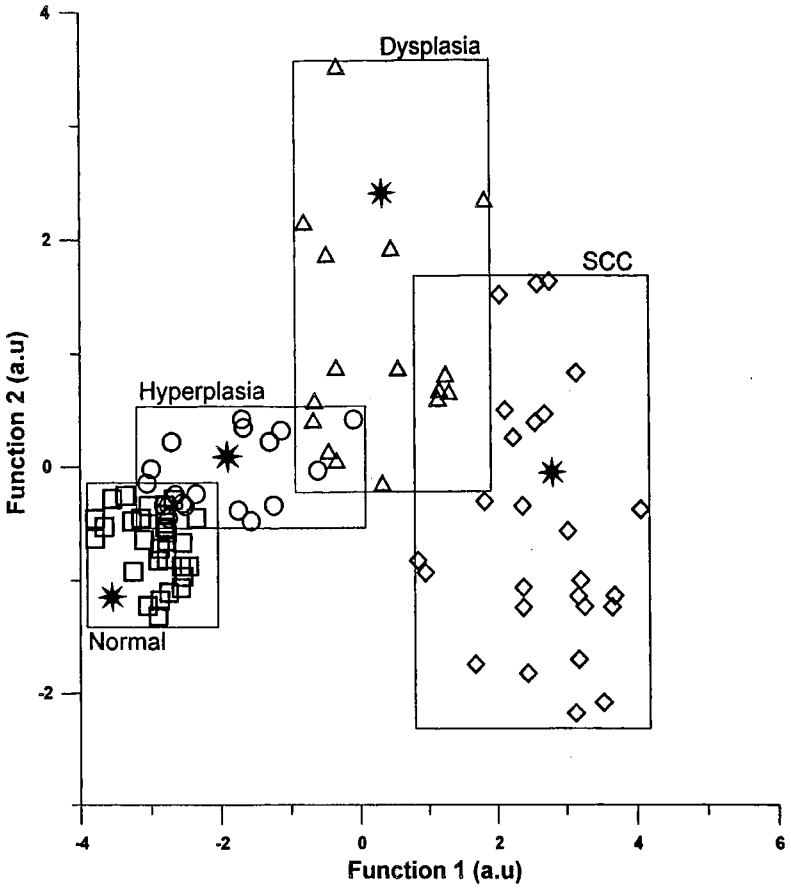

Fig. 4.4. Scatter plot of the first two discriminant functions by LDA. The four categories of oral tissues are located in the four distinct areas. The results are presented according to the discriminant function scores determined based on LOO method of cross validation.

Earlier studies have shown that measurement from healthy tissues within a patient is not well defined, due to 'field cancerization' and influence of carcinogens like tobacco, pan and alcohol (Slaughter et al., 1953; Wang et al., 2003a; de Veld et al., 2003; Mallia et al., 2008a). Therefore, the present study has relied on spectral measurements from healthy population as control and attained improved sensitivities and specificities for tissue discrimination (Table 4.1). Nevertheless sites such as vermilliion border (VB) of the lip,

dorsal and lateral sides of the tongue could not be studied due to the presence of porphyrin / bacteria emission at 635, 685 and 705 nm (Mallia etal., 2008a; de Veld et al., 2003).

While different methods were used for preprocessing of spectral data, normalization with respect to the autofluorescence peak intensity was found to give better results in classification. In comparison, mean-scaling of spectral data and mean-scaling of normalized data did not improve the classification efficiency.

First three PC loadings (Fig. 4.2) of the spectral data set with laser excitation at 404 nm resembled the average autofluorescence spectrum (PC1), while oxygenated hemoglobin absorption dips and porphyrin like peaks were observed in PC2 and PC3. The three PCs together form 98.6% of variance in the spectra, which means that the complete spectral data set could well be described by these three PCs. Significant differences observed in the mean PC scores (p-value < 0.005) between different lesions, with low standard deviation, shows that tissue classification is possible with the information contained in these PCs. The shape of the PC loading is responsible for the observed significant differences.

An overall sensitivity of 100%, 98% and 100% respectively was obtained in this clinical trial, for discriminating normal from benign, benign from premalignant and premalignant from malignant tissues with corresponding specificities of 98%, 100% and 95% (Table 4. 1). Using partial least-squares and artificial neural network (PLS-ANN) classification

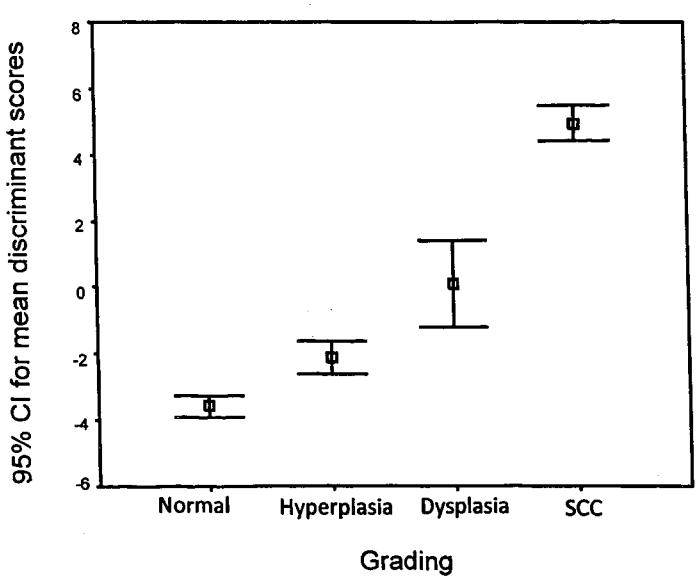

Fig. 4.5 Discriminant function scores plotted with their error bars for different lesions at 95% confidence interval for the mean.

algorithm, Wang *et al* (2003a) obtained a sensitivity of 81% and a specificity of 96% for discrimination of premalignant and malignant from benign tissues. Similarly, a sensitivity of 92% and a specificity of 95% was obtained for discriminating benign from dysplasia or SCC using partial least squares discriminant analysis (PLSDA) of autofluorescence spectral data for *in vivo* diagnosis of hamster buccal pouch pre-cancers and cancers (Wang et al., 2003b). In another study, de Veld *et al* (2005) have reported an improved sensitivity of 94% and a specificity of 94% for distinguishing cancerous lesions from normal. As compared to all these studies, the diagnostic accuracies obtained in this clinical trial are vastly improved and could be attributed to the use of healthy volunteer data as control and improvements in the classification algorithm.

For group classification, LDA is a well established method, which unites input parameters into a discriminant function to classify the available data into different groups (Dzendrowskyj et al., 2001; Fernandez et al., 2002). In this study PCA was performed only for the reduction of spectral intensity data. Tissue classification was made purely based on the discriminant function scores obtained by LDA for each lesion. Cut-off values in the scatter plot were used to calculate diagnostic accuracies. The results obtained show the ability of this procedure to act as an adjuvant to clinicians for *in vivo* tissue differentiation in real-time without any tissue removal.

The discriminant function scatter plot is shown in Fig. 4.4 for all the four groups. In this figure the first two functions in LDA classifies the lesions into four different groups based on group centroids. Group means are centroids and differences in location of centroids represent dimensions along which the groups differ. It is possible to visualize discrimination between groups by plotting the individual scores for the two discriminant functions. This classification based on the Mahalanobis distance is a multivariate measure of the separation of a point from the mean of a dataset in n-dimensional space. The sample is classified to the group from which it has shorter Mahalanobis distance. All the normal and SCC lesions

are classified correctly in Fig. 4 while there is an overlap in the case of hyperplasia and dysplasia samples. One dysplasia site was misclassified as SCC, whereas one hyperplasia was misclassified as normal and another one as dysplasia. Rectangular boxes are drawn to demarcate each group. The error bars of the mean discriminant function scores at 95% confidence interval are plotted in Fig. 4.5. Lack of overlap between different groups shows the statistical significance of the study result and confirms the classification potential of the mean discriminant function scores. The results presented reveal the potential of this method to accurately discriminate different lesion types using discriminant function scores.

4.5. Conclusions

The application of LDA-LOO method on the autofluorescence spectra recorded during this clinical trial was suitable to discriminate oral mucosal alterations during tissue transformation towards malignancy with improved diagnostic accuracies. The present study was able to classify dysplasia from SCC, dysplasia from hyperplasia and hyperplasia from normal with sensitivities of 100%, 98% and 100%, respectively with corresponding specificities of 98%, 100% and 95%. The results obtained show improved sensitivity and specificity as compared to previous reports and confirm the advantages of using multivariate statistical analysis on LIAF spectral data for non-invasive diagnosis of oral pre-malignancies. Further measurements are envisaged in a larger population group to explore the applicability of discriminant functions in grading of oral mucosa in real-time and to improve the diagnostic accuracies for discrimination of hyperplastic and dysplastic lesions from normal and to distinguish OSF from other pre-malignant conditions.