Human Saliva as a Substitute Diagnostic Medium for the Detection of Oral Lesions Using the Stokes Shift Spectroscopy: Discrimination among the Groups by Multivariate Analysis Methods

Pavan Kumar\textsuperscript{1,2*}, Asima Pradhan\textsuperscript{2}

Abstract

Objective: Our objective in the present study is to detect oral mucosal lesions non-invasively by probing two solutions with reference to diagnostic technique and non-invasive media. In the diagnostic technique, Stokes shift (SS) spectroscopy (SSS) has been utilized for the detection of oral lesions. In the diagnostic media, human oral tissue and saliva are included. Methods: SS measurements are carried out on oral squamous cell carcinoma (OSCC), dysplastic (precancer), and normal/control tissue and saliva samples. Measurements are performed on 86 tissue and 86 saliva samples using the commercially available spectrofluorometer. Offset wavelength of 120 nm, which is the Stoeke shift of nicotinamide adenine dinucleotide (NADH) has been selected over the other offsets (i.e., 20, 40, 70 and 90 nm). Result: Presence of tryptophan, collagen, NADH, and flavin adenine dinucleotide (FAD) bands were noticed in the SS spectra of tissue. Like the tissue spectra, presence of these bands was also found in the SS spectra of saliva except the collagen band. Classification among the samples accomplished by the make use of multivariate analysis methods. In the multivariate analysis methods, principal component analysis (PCA) is applied first on SS data of tissue and saliva and then Mahalanobis distance (MD) model and receiver operating characteristic (ROC) analysis employed successively. Overall accuracy values of 94.91 %, 84.61 %, and 85.24 % were obtained among OSCC to normal, dysplasia to normal, and OSCC to dysplasia for tissue samples and 88.46 %, 90.16 % and 94.91 % accuracy values were obtained for saliva using the SS spectroscopy. Conclusion: Obtained results of human saliva are equivalent to human oral tissue using the SS spectroscopy. It indicates that saliva may be utilized as a substitute diagnostic medium and SS spectroscopy as a diagnostic technique for non-invasive detection of oral lesions at the primarily stage.

Keywords: Stokes shift spectroscopy- oral cancer- human tissue- human saliva- multivariate analysis

Introduction

A significant progressive increase of oral mucosal lesions in Asian countries is a major concern to everyone. Worldwide, among all the cancers, oral cancer is the eighth most common and one of topmost among males and females in India (Coelho, 2012; Ferlay et al., 2015). Absence of early symptoms and late diagnosis are two major causes of higher morbidity and mortality rates. To overcome the incident and death rates, it is needed to detect and diagnose cancer at initial stage. Studied have revealed that early diagnosis can significantly enhance the chances of cure. It is therefore need to find some novel solutions in terms of diagnostic media and diagnostic techniques.

Visual inspection, brush biopsy, toluidine blue, histopathology etc., are some of conventional techniques to detect the oral lesions. Among the conventional techniques, tissue biopsy with histopathological examination is the gold standard tool for the detection oral cavity lesions (Scully, 2008; Patton et al., 2008; Omar, 2015). The major complication with this technique is its invasive and time-consuming process. All Patients having the abnormality either it is benign, dysplastic (mild, moderate, severe), or OSCC have to go through this painful process during the treatment. Sometimes, patients have to go through several biopsies to identify the right abnormal area of tissue. To overcome these complications, there is a need of such tools which can examine the lesions non-invasively. Several non-invasive techniques such as fluorescence spectroscopy and imaging, Raman spectroscopy and imaging, diffuse reflectance (DR) spectroscopy and imaging have been adopted by many

\textsuperscript{1}Department of BSH, Faculty of Engineering and Technology (FEAT), Datta Meghe Institute of Higher Education and Research (DMIHER), DU, Sawangli - 442001, India. \textsuperscript{2}Department of Physics, Indian Institute of Technology (IIT) Kanpur, Kanpur-208016, India. *For Correspondence: pavan2012iitk@gmail.com
research groups along with the clinician to identify the oral mucosal lesions (Ramanujam et al., 1996; Pichardo et al., 2007; Majumder et al., 2008; DeCoro et al., 2010; Alfano, 2012; Singh et al., 2016, Kumar et al., 2023). Fluorescence based techniques have been widely utilized by the research groups for the identification of oral lesions (De Veld et al., 2005; Lane, 2006; Tsai et al., 2008; Singh et al., 2012; Kumar et al., 2019, Sah et al., 2023).

Stokes shift (SS) spectroscopy (SSS), which is also referred to as synchronous fluorescence (SF) spectroscopy (SFS) has been used by few research groups for cancer detection (oral cancer, breast cancer and cervical cancer) (Vo-Dinh et al., 1978; Majumder et al., 2000; Alfano and Yand, 2003; Devi et al., 2015). J Ebenezar et al. group used SF spectroscopy on normal and abnormal cervical tissues samples and found multiple non overlapping fluorophore bands in a single scan. They were able to discriminate the lesions with the accuracy of 100% (Ebenezar et al., 2010). Yang Pu et al group used SS spectroscopy (Δλ = 40 nm) on cancerous and normal breast tissues. Multiple distinct bands of tryptophan, collagen and NADH were found in SS spectra. By analyzing the data, they were able to differentiate the cancerous tissues from normal with the accuracy of 83.33% (Pu et al., 2012). Many research groups have studied human saliva for the detection of oral lesions using the various techniques but its detection using the spectroscopic devices are still need to execute in large scale.

It is established that biochemical and morphological changes occur with the progress of disease in the tissue of oral cavity. Biochemical changes are also observed in human saliva. Collagen, tryptophan, flavin adenine dinucleotide (FAD), NADH, and porphyrins are some of the key fluorophores present in the oral tissue and alteration in the concentration of these fluorophores are observed with the progress of disease. Except the collagen, fluorophores such as NADH, tryptophan, FAD, DNA and RNA, are also found in human saliva and variation in the concentration also occur with progress of cancer. Among the body fluids, human saliva can be best choice for the detection of oral lesions due to its non-invasive collection process as well as direct contact of fallen cells in the oral cavity. Its extensive production (~1 to 1.5 liters) per day in an adult human and ease in sample collection makes it potential campaigner. Several studies have been performed by the research groups over the last 20 to 25 years for various purposes (Markopopos et al., 2010; Wu et al., 2010; Pfaffe et al., 2011; Chenge et al., 2014; Kuznetsoy et al., 2015). It has been studied in forensic cases such as sexual assault, child abuse, as well as utilized to detect HIV infected patients (Soulos et al., 2000; Vinkler et al., 2009). In the cancer diagnosis study, it has been utilized for the detection of breast cancerous lesions (Hosseini et al., 2009). In couple of years, human saliva is widely tested for the novel coronavirus detection (To et al., 2019). To detect the lung cancer using saliva, a study has been executed by the research group on cancerous patients and normal volunteers and accuracy of 80% are achieved (Xiaozhou et al., 2012). For oral cancer detection on cancerous and control groups using the fluorescence spectroscopy on human saliva, sensitivity of 85.7% and specificity of 93.3% were achieved (Yuvraj et al., 2014). LIFS technique used on saliva samples by the research group for the detection of oral cancer and achieved a sensitivity and specificity values of 79 % and 78 % respectively (Patil et al., 2013). A concentration-based study on oral carcinoma and normal saliva samples are conducted and the sensitivity and specificity values of 71 and 75 percent respectively are found (Nager et al., 2006). A study conducted on human saliva by our group on patients (OSCC & OSMF) and volunteers for oral cancer detection reported 93 %, 95 % and 92 % accuracy values (Kumar et al., 2018). A study conducted on human saliva for the detection of head and neck cancer using fluorescence spectroscopy, three groups i.e., SCC, dysplastic and normal were differentiated with the accuracy values of 98 %, 93 % and 81 % (Kumar, 2022).

As a background study, we have chosen different offset wavelengths i.e., 20, 40, 70, 90 and 120 nm. Among the offsets, Δλ = 120 nm on tissue samples captures the maximum fluorescent bands of collagen, tryptophan, NADH, FAD, and porphyrin with differences in fluorescent intensities among the three groups. Δλ = 70 nm on tissue samples has captured fluorescent bands of collagen, tryptophan, NADH, and FAD and displayed difference in intensities among the bands. In human saliva at Δλ = 120 nm, though most of the captured fluorescent bands display differences among the groups. However, Δλ = 70 nm mainly captures only one major band of tryptophan and minor bands of NADH and FAD and does not display significant differences. Other offsets did not display as many bands as 70 and 120 nm have shown and consequently differences among the groups. In the systematic data analysis using multivariate methods i.e., PCA, Mahalanobis, and ROC for 70 and 120 nm Δλ values, 120 nm yielded the maximal classification among the three groups in both media (tissue and saliva). Therefore, in this present paper, we have included the result of 120 nm SS value only.

Materials and Methods

Sample Collection

Three groups of patient/volunteers i.e., OSCC, dysplastic and normal are included in the study and sample collection was done. In these three groups, 86 tissue samples (OSCC=34, dysplastic = 27 and normal = 25) and 86 saliva samples (OSCC = 34, dysplastic = 27 and normal = 25) were collected. Normal tissue samples were taken from the unaffected area of OSCC and dysplastic patients. However, saliva samples were taken from twenty-five control group. Control group included in this study confirmed that they do not take any tobacco-based products. Significant variation in the patient age were noticed. OSCC and dysplastic patients spanned between 34-85 (average age with the SD 47±13) and 22-65 (average age with the SD 41±15) respectively. Control group was 25-56 (average age with SD 36±10). Patients reported to clinicians during their first appointment for the treatment, were instructed that they would not consume any food items on the day of biopsy. Saliva collection were done in small sterile containers (5 ml) which were given to patients.
and instructed to produce the saliva at least 1 ml. Once the samples were collected, patients were informed to go for biopsy. After completion of biopsy, tissue samples were collected in a sterile. Tissue and saliva samples acquisitions were done in Hallet hospital. Hallet hospital is affiliated to G.S.V.M. Medical College, Kanpur, India. An ice box was used to bring these samples to IIT Kanpur campus. Prior to start the experiment, tissue sample was cleaned in the saline water. Further, sample was mounted on a quartz cuvette and SS measurements were performed. Like the SS measurements on tissue, saliva sample was first poured into a cuvette and spectra were recorded. After conducting the measurements, tissue samples were sent for histopathology. Histopathology reports of the patients were received within the 10 to 15 days and compared with the SSS results.

**Measurement Techniques**

Stokes shift (SS) spectra of the samples were recorded on a spectrofluorimeter setup (Fluorolog 3, Model FL3-22). Fluorolog 3 setup is equipped with an excitation and emission monochromators, a Xeon source and other accessories. During the measurement on tissue and saliva samples, slit widths of both monochromators were adjusted at 1 nm and SS spectra were recorded. These spectra were recorded at 0.1 integration time in the scan range of 250 to 600 nm.

**Analysis Methods**

Principal component analysis (PCA) is applied on the recorded SS data sets of tissue and saliva in the spectral range of 250 to 600 nm. It is a statistical method and adopted to convert a higher dimensional data set into a lower dimensional data set of uncorrelated variables (Abdi and Williams, 2010). Principal components (PCs) of correlation matrix are calculated which are also known as eigen vectors and then PC scores are estimated. Classification among the groups is accomplished by using Mahalanobis distances (MD) model. In the MD model, firstly PC scores are segregated into training data (t) and validation data (V) sets also called as known and unknown groups respectively and then Mahalanobis distance (D_maha) within and among the three groups are computed with the help of following equation as shown below known as Mahalanobis distance.

\[
D_{\text{maha}} = \sqrt{(V - \mu_{t})^T C_t^{-1}(V - \mu_{t})}
\]

Here \(\mu_{t}\) is the mean of training data and and \(C_t^{-1}\) denotes inverse of the correlation matrix. After finding the MD values, receiver operating characteristic (ROC) analysis is employed on these values (binary data set at a time) and sensitivity, specificity, and accuracy values among the samples are estimated (Kumar et al., 2019; Akobeng, 2007).

**Results**

**Stokes Shift Spectroscopy (SSS) for Human Oral Tissue**

Area normalized averaged spectra of OSCC (n=34), dysplastic (n=27), and normal (n=25) oral tissue samples at SS of 120 nm in the spectral range of 250 – 600 nm is shown in Figure 1(a). SS spectra display a band around 282 nm with two sharp peaks at 280 and 291 nm, which are due to tryptophan and collagen respectively and can be seen more clearly in inset of Figure 1(a). Another broad band of NADH is observed, which is peaked near 347 nm. Dips around 286 and 309 nm are observed, which are attributed to absorption by porphyrin and blood respectively. A minor band near 438 nm is due to FAD and other bands near 515, 551, and 582 nm are due to porphyrins. These minor bands of FAD and porphyrin can be seen clearly in the typical spectra shown in Figure 1(b).

**Stokes Shift Spectroscopy for Human Oral Saliva**

Area normalized averaged SS spectra of OSCC, dysplastic (n=27), and normal (n=25) oral saliva samples at SS of 120 nm in the spectral range of 250 – 600 nm are displayed in Figure 2(a). In the SS spectra of saliva, one can observed two major bands attributed to tryptophan and NADH near 270 and 340 nm respectively. The tryptophan band does not have as much difference in intensity among the groups as NADH. Mean peak positions of NADH bands with the standard deviation (SD) for OSCC, dysplastic, and control groups are at 340±7nm, 344±7.3 nm and 347±9.8 nm. A dip near 286 nm is observed,
Figure 2. SS Spectra of OCCC, Dysplastic, and Normal Saliva Samples (a) area normalized averaged spectra (b) typical spectra

Figure 3. Scatter Data Plots of First Three PC Scores for (a) tissue samples (b) saliva samples

as seen clearly in inset of Figure 2 (a), is attributed to absorption by porphyrin. It is not as significant as observed in SS spectra of tissue samples. In the averaged spectra, minor bands of FAD (425 nm) and porphyrins (503 and 513 nm) cannot be seen. However, presence of these minor bands is observed in few samples. Typical spectra

Figure 4. Scatter Plots of Mahalanobis Distances for (a) Normal to OSCC tissue samples (b) Normal to Dysplasia tissue samples (c) Dysplasia to OSCC tissue samples (d) ROC curves
displaying the minor bands are shown in Figure 2 (b). FAD bands are more commonly seen in OSCC and dysplastic groups. However, porphyrin bands are seen only in few cases of OSCC and dysplastic groups.

**Discussion**

As discussed in the Analysis Methods subsection, PCA was employed in the entire spectral range of the SS spectra to reduce dimension. First seven PCs (PC1 to PC7) almost covers variance of ≥ 99% were taken for the data sets of tissue and saliva. First three PC scores of the SS data of oral tissue and saliva which captures the variances of 92% and 89% respectively are shown in Figure 3(a) and (b) respectively. In Figure 3(a), it can be seen that the clusters formed by the normal group is well separated from the overlapped clusters formed by dysplastic and OSCC groups. Like it, scatter plot of first three PC scores procure from SS spectra of saliva are presented in Figure 3(b). In it, the cluster of the dysplastic group overlaps on both sides with the normal and OSCC groups. PC score of normal samples (tissue and saliva) are localized compared to the OSCC and dysplastic samples in both media. It indicates that there is more variability in the spectral data of OSCC and dysplastic groups.

After estimating Principal components, PC scores of both the data sets are computed and loaded in MD model to find the MD values. Calculation of MD values were done on training and validation sets of PC scores. Plots of MD values among the groups and vice versa of the SS data of tissue are plotted against the total number of samples and displayed in Figure 4a-c. ROC employed on the MD values of tissue samples differentiate OSCC to normal, dysplasia to normal, OSCC to dysplasia with sensitivities of 94.12 % (32/34), 88.89 % (24/27), 88.24 % (30/34) and specificities 96% (24/25), 80% (20/25), 81.48% (22/27) with the overall accuracies of 94.91 % (56/59), 84.61 % (44/52) and 85.24 % (52/61), respectively. Likewise, Mahalanobis distances of saliva data are calculated and plotted against the number of samples. These distances are displayed in Figure 5a-c. ROC on the MD distances of saliva samples differentiate the respective groups with sensitivities of 94.12 % (32/34), 81.48 % (22/27), 91.17 % (31/34) and specificities 96 % (24/25), 96 % (24/25), 88.89 % (24/27) with the overall accuracies of 94.91% (56/59), 88.46% (46/52) and 90.16 (55/61), respectively. ROC curves of tissue and saliva are shown in Figure 4(d) and Figure 5(d) respectively. While computing the MD values, training and validation data sets are chosen randomly several times from each group and calculations are done but significant change in sensitivity and specificity values are not noticed. Scatter plots between OSCC to dysplasia show larger overlap than the OSCC to normal and dysplasia to normal groups on tissue as well as in saliva. However, the sensitivity and specificity values are almost comparable. It shows that confidence level for discrimination OSCC to dysplasia is lower than the other two cases.

In conclusion, a comparative study carried out for the detection of oral cancer using SS spectroscopy at Δλ value of 120 nm on tissue and saliva samples captured several distinct bands in a single scan. In the SS spectra of tissue, major bands of tryptophan, collagen, and NADH were observed while in saliva contributions were mostly from tryptophan and NADH bands. Minor bands of FAD and porphyrin were found in both the media. The first two overlapped peaks at 280 and 291 nm in tissue spectra were due to tryptophan and collagen respectively while
the first band near 270 nm in saliva was due to tryptophan. Collagen band and blood absorption dip were absent in saliva, as expected. Since the collagen is a structural protein which is only found in tissue. To extract the information from the SS spectra, PCA was employed first. Further MD model was used for classification among the three groups and then ROC was applied to compute the diagnostic parameters. ROC on the Mahalanobis distance obtained from PC scores of SS spectra of oral tissue and saliva was able to differentiate among the groups with sensitivities ranging from 94 to 81 percent and specificities from 96 to 81 percent. Results on human saliva were as promising as those from oral tissue using MD model. The study thus reveals that human saliva may be used as an alternate medium and SS spectroscopy is seen to be an excellent tool for the detection of oral cancer at an early stage.

Author Contribution Statement

Measurements, data analysis, and first paper draft were completed by corresponding author and further corrections were done by both the authors.

Acknowledgements

Author would like to acknowledge to Dr. Surendra Kumar Kanaujia, who is an ENT specialist at GSVM Medical College Kanpur. He helped us in providing the samples. This research work was possible to conduct with the consent of patients and volunteers. We all want their good health.

Funding Statement

To perform the presented research work, facilities of bio photonics lab (SL-111) in the Physics department at IIT Kanpur are utilized.

Approval

Presented research for the publication is my thesis work. We the authors are giving approval for publication of this work.

Ethical Declaration

Prior to start this clinical work, we had taken the approval from the committee members of IIT Kanpur and GSVM medical college, which is according to the declaration of Helsinki.

Conflict of Interest

There is no conflict of interest.

References


This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.