L.4: Anatomical variability of *in vivo* Raman spectra of normal oral cavity and its effect on oral tissue classification

An important factor that is expected to affect the outcome of *in vivo* Raman spectroscopic diagnosis of oral cancer is the large intra-patient as well as inter-patient variability in the intensity and line shape of the measured tissue Raman spectra arising from the anatomical differences over the different tissue sites interrogated. A clinical *in vivo* study was carried out at Tata Memorial Hospital (TMH), Mumbai on healthy volunteers and patients with various oral lesions to address this issue. The objectives were two-fold: (i) to characterize the variability of the *in vivo* Raman spectra of the different anatomical sites of the oral cavity of healthy volunteers, and (ii) to investigate the effect of this inter-anatomical spectral variability on the performance of the diagnostic algorithm developed to discriminate malignant and potentially malignant oral lesions from the healthy oral mucosa.

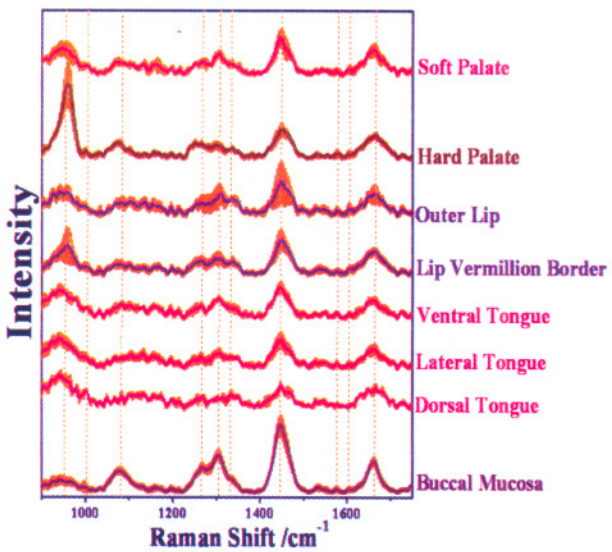

Fig. L.4.1: Mean, normalized Raman spectra of the different anatomical sites of oral cavity of healthy volunteers.

An unsupervised cluster analysis using Fuzzy c-means clustering algorithm was conducted for quantifying the underlying structure of the normal oral tissue spectra. The algorithm was found to segment the normal oral tissue sites, based on similarity of spectral patterns, into four major anatomical clusters (AC): (1) outer lip, and lip vermillion border into AC-I with an accuracy of 80%; (2) buccal mucosa into AC-II with an accuracy of 72%; (3) hard palate into AC-III with an accuracy of 92%; (4) dorsal, lateral and ventral tongue and soft palate into AC-IV with an accuracy of 76%.

A probabilistic multi-class diagnostic algorithm, developed for supervised classification, was used to classify the whole set of measured tissue Raman spectra into three categories: normal, potentially malignant and malignant. The results listed in Tables - L.4.1 and L.4.2 show that the diagnostic algorithm, when applied on the pooled set of spectra from all the clusters, correctly discriminates normal, malignant and potentially malignant tissue sites with 86%, 88%, and 86% accuracy respectively, which amounts to an overall accuracy of 87%. However, when the anatomy-matched data sets are considered, the overall classification accuracy is found to improve to 95% with the algorithm correctly discriminating the corresponding tissue sites with 94%, 99%, and 91% accuracy respectively.

Table L.4.1: Confusion matrix displaying results of classification of the Raman spectra of oral tissue sites without anatomical clustering.

Pathology Diagnosis	Raman diagnosis		
	N	M	PM
Normal (N)	86%	5%	9%
Malignant (M)	7%	88%	5%
Potentially Malignant (PM)	7 %	7%	86%

Table L.4.2: Confusion matrix displaying results of classification of Raman spectra of anatomy matched oral tissue sites.

Pathology Diagnosis	Raman diagnosis		
	N	M	PM
Normal (N)	94%	2%	4%
Malignant (M)	0%	99%	1%
Potentially Malignant (PM)	3%	6%	91%

Reported by:
H. Krishna and S. K. Majumder (shkm@rrcat.gov.in)