RSC Advances

ARTICLE

Received 00th January 20xx, Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Global Spectral and Local Molecular Connects for Optical Coherence Tomography Features to Classify Oral Lesions towards Unravelling Quantitative Imaging Biomarkers

Satarupa Banerjee,[a] Swarnadip Chatterjee,[b] Anji Anura,[a] Jitamanyu Chakrabarty,[c] Mousumi Pal,[d] Bhaskar Ghosh,[e] Ranjan Rashmi Paul,[d] Debdoot Sheet[f] and Jyotirmoy Chatterjee[a]

Biopsy based diagnosis of oral precancers like leukoplakia (OLK) and submucous fibrosis (OSF) as well as squamous cell carcinoma (OSCC) suffers from observer specific variability. Present work explored the utility of intensity and textural features from optical coherence tomography (OCT) images after specific feature subset selection for precise classification of oral lesions using variants of support vector machine. Concomitant application of Fourier transform infrared (FTIR) spectroscopy for endorsing global biochemical signatures, and histochemistry was performed further for value addition of the OCT findings. Immunohistochemical findings for characterization of specific local molecular alteration were also included in this. Result suggested that, OCT features could differentiate the lesions with high sensitivity and specificity. The FTIR result showed glycogen, keratin and carbohydrate related alteration in OSCC, decrease in collagen specific amino acids and skeletal muscle related proteins in OSF and distinct variation in tissue hydration status in diseases. There was also increase in keratin layer thickness in OLK due to overexpression of Cytokeratin10 in superficial layer; while in OSF, skeletal muscle was found to be replaced with dense collagen I. These disease specific alterations were assumed to be the underlying phenomenon associated with intensity and textural variations in OCT images, using which specific quantitative imaging biomarkers were proposed.

Introduction

Optical diagnostic systems like optical coherence tomography (OCT), Fourier transform infrared spectroscopy (FTIR), Raman spectroscopy, microendoscopy, and fluorescence spectroscopy are effectively emerging for non-invasive studies of pathologies, especially for characterization of pre-cancer and cancer. These techniques also help in value addition to the existing histopathological diagnostic gold standard as well as molecular pathology towards exploration of newer information.[1]

OCT, a non-invasive imaging technique, provides real-time, high-resolution, micro-architectural sub-surface images of nearly up to 2mm tissue depth. [2] Previous studies correlated healing progression and maturation of epithelial and sub-epithelial components considering OCT image attributes and histopathological features.[3] 'Lucidity' is the optical intensity descriptor used for interpreting OCT images. It tends to vary in different regions of layered body structure like oral mucosa, skin wounds etc. Since the operating principle of OCT imaging is governed by backscattering of light and exploiting a 'biological window' with minimal absorption, the changes in tissue refractive index modulate intensity characteristics.[4] Such scattering also depends upon tissue structural components, surface roughness[5], hydration cum maturation status, nuclei size, presence of collagen fibres, keratin content[6], tissue type[7] and membrane lipid density of cells[8]. In skin[9], cervix[10] and oral mucosa[11] transition zones and architectural changes during disease progression can be identified by OCT. Such demarcations are possible due to differential thickness and composition of epithelial or sub-epithelial layers.[12, 13] In this context, Ughi et. al utilized intravascular coronary OCT for differentiating normal and abnormal pathologic condition by textural image analysis.[14] A recent study also implemented automated classification of oral malignancy in hamster buccal pouch model using OCT textural features.[15] However, further scopes are there to enhance the diagnostic efficiency of such OCT images for oral mucosal lesions, by corroborating chemical and molecular signatures of tissues documented by FTIR, histochemistry (HC) and immunohistochemistry (IHC). The present study therefore primarily delved to classify oral lesions

a. School of Medical Science and Technology, Indian Institute of Technology, Kharagpur, India
b. Advanced Technology Development Centre, Indian Institute of Technology, Kharagpur, India
c. Department of Chemistry, National Institute of Technology Durgapur, India
d. Department of Oral and Maxillofacial Pathology, Guru Nanak Institute of Dental Science and Research, Kolkata, India.
e. Department of ENT & Head neck Surgery, Medical College, Kolkata, Kolkata, India
f. Department of Electrical Engineering, Indian Institute of Technology, Kharagpur, India Address here.
Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

on the basis of intensity and textural features of OCT images, besides providing tissue architectural information, and also to amalgamate information obtained from other modalities like FTIR and HC/IHC towards better characterization of oral lesions. The FT-IR and HC / IHC are considered for global assessment of biochemical variation and local composition/gene expressional changes respectively.

FTIR is a widely used low cost tool for chemical portrayal of materials, yet underexplored diagnostic modality for spectral characterization of biopsied tissues.[6] In this perspective, FTIR in transmission mode was used for functional group analysis and disease specific chemical characterization of oral lesions from global dimension. HC and IHC findings also provided local specific compositional alteration. Periodic acid–Schiff (PAS) depicted information on polysaccharides as well as keratins and Van Gieson's (VG) staining illustrated differential staining of collagen and other connective tissue components[16]. The IHC study of collagen I (COL-I) and cytokeratin 10 (CK 10) expressions endorsed the vital compositional[17] and maturational[18] information respectively and corroborated with the tissue architecture.

After analyzing the same tissues under different modalities, viz. OCT, FTIR, HC and IHC, two propositions were considered. Firstly, oral lesions can be segregated on the basis of a specific subset of intensity and textural features extracted from OCT images which could be further proposed for optimum disease segregation. Recently quantitative imaging biomarkers (QIBs) are defined as "an imaged characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or a response to a therapeutic intervention".[17] The concepts of QIBs further helped to assume that, if biochemical characterization of the same tissue sections can be performed, then selected OCT features could be rechristened to QIBs. Therefore support vector machine (SVM) was used here for disease classification, since it can classify the diseases with high predictive accuracy, medium fitting speed, and good prediction speed along with memory as shown in previous studies.[18, 19] Quadratic and cubic kernels were also used here to manipulate the efficiency of the learners, since they are commonly used non-linear kernel beside linear one.[20] After feature reduction using minimum Redundancy Maximum Relevance (mRMR) algorithm[21] for feature subset identification followed by the classification task and biochemical characterization of the tissues using specified modalities, QIBs were thus proposed.

Secondly, it was assumed that disease specific difference in the textural feature was due to disease specific changes in biochemical component at tissue level. The uniqueness in the present work is not only providing structural information but to treat OCT as a measurement modality which cannot be interpreted by a human observer. This may overcome the limitation of need of expert based disease diagnosis. It was also assumed that multimodal approach may provide complementary information, where disease specific difference in the intensity and textural features of OCT can also be logically correlated with characteristic molecular pathology attributes. Therefore underlying chemical alterations were also sought to validate the notion that difference in the global chemical signatures in different disease condition may be associated with changes in the intensity and textural features. In previous studies, amalgamation of the morphological information of OCT and biochemical information for diagnosis of diseases resulted in increased specificity and sensitivity,[22] whereas this study was performed in fragmented manner to highlight utility of each modality.

Two pre-cancers (viz. oral leukoplakia (OLK) and oral submucous fibrosis (OSF)), beside oral squamous cell carcinoma (OSCC) were chosen in this study. OLK is presented by white plaques of questionable risk having excluded from other known diseases or disorders that carry no increased risk for cancer,[23] whereas OSF, defined as a chronic, premalignant condition is characterized by progressive sub-epithelial fibrosis.[24] The reasons behind selection of the two lesions are, their high malignant potentiality and despite having differences in their origin, they both culminate into OSCC.[19] OSCC is defined as "a malignant epithelial neoplasm exhibiting squamous differentiation as characterized by the formation of keratin and/or the presence of intercellular bridges".[25] It may also be emphasized that, consumption of tobacco (smoked or smokeless) and areca nut are the major risk factors associated with OLK and OSF respectively.[23]

Finally on the basis of textural and intensity attribute selection in OCT images, molecular characterization of tissues with the OLK, OSF and OSCC as well as logical integration of the results, QIBs could be proposed, which is the main aim of this study. Multimodal diagnostic evaluation of oral lesions in turn thus not only addressed diagnostic ambiguity, but also emphasized role of value addition in translational research towards better disease characterization.

Materials and Method

Sample collection, OCT imaging and Tissue processing: In vivo OCT images were acquired from selected 57 patients (Age 18-65). Clinical diagnostic criteria of the diseases were provided in Supplementary Table 1. Biopsy samples (7 Normal (NOM), 11 OSF, 16 OLK and 23 OSCC) were also collected by the oncopathologists from the same area of oral cavity of the patients in GNIDSR, Kolkata under ethical clearance of institution ethical committee (GNIDSR/IEC/ECC/2015/010 dt. 08/01/2015). Informed consent was obtained from all the subjects (both normal and diseased) recruited in the study. During NOM patient selection, only age and sex matched subjects were considered for this study. The tissue biopsies were fixed in 10% formaldehyde solution in phosphate buffered saline.

OCT imaging and tissue processing: In vitro preserved biopsy samples were subjected to SS-OCT imaging (Model: OCS1300SS, Thorlabs-Inc., Newton, NJ, USA having scanning pulsed laser with center wavelength of 1,325 nm, half-power spectral bandwidth < 100 nm, axial scan rate of 16 kHz, coherence length of 6.0 mm, and average output power of 10 mW) using the method of Sheet et al.[11] 3D image volumes of the whole tissue were acquired and subsequently 2D images with presence of disease area as suggested by onco-pathologists were used in this study. Image resolution of each 2D transverse OCT scans were 512 × 512 pixels corresponding to 3 mm×3 mm physical size of the imaged section. 16 NOM, 41 OLK, 51 OSF and 64 OSCC OCT images were considered in the study.

After OCT imaging of fixed tissues, they were paraffin embedded. 4 µm thick sections were mounted on six albumin coated glass slides and two poly-L-Lysine coated glass slides. All the sections were de-paraffinized using xylene. Albumin coated slides were used for FTIR data acquisition, H&E (Haematoxyline and Eosine) staining based histology, and HC, while the poly-L-Lysine coated slides were used for IHC staining.

Histological, histochemical and immunohistochemical staining: The tissue sections were placed on albumin coated glass slides. During H&E staining, the sections were stained with Harris' hematoxyline and counter stained with eosin. PAS staining was performed for glycogen and VG for collagen[16]. Briefly, in PAS staining deparaffinized tissue sections were first oxidized with 0.5% Periodic acid for 10 minute, stained with Schiff reagent for 5 minute and counter stained with Harris' hematoxyline. During VG staining, the deparaffinized samples were stained with Harris' hematoxyline and counterstained with picric acid and acid fuschin (9:1) mixture. Tissue sections were baked for 30 minutes at 60°C followed by deparaffinization, gradual alcohol hydration for IHC staining, and then subjected to antigen retrieval (EZ-Retriever System V.2; BioGenex, USA) in 10 mM tris-ethylenediaminetetraacetic acid buffer (pH 9.0). Chromogenic methods were used to immunostain the sections. Staining for CK10 and COL-I was performed. Immuno-detection was performed with Horseradish Peroxidase conjugated secondary antibody with chromogen 3, 3'-diaminobenzidine and counterstained with Harris' Hematoxyline using Super Sensitive Polymer-HRP IHC Detection System kit. The images were grabbed digitally under 20× objectives using a bright field inverted microscope (Zeiss Observer. Z1, Carl Zeiss, Germany).

Feature extraction from OCT image: OCT images were segmented using 'Image Segmenter' app of MATLAB 2015a version. Initialization of the area was manually performed, on which edge based active contour segmentation was implemented and evolved with 100 iterations. Then eight intensity and 14 textural features were extracted from the semi-automatically segmented area of both epithelium and sub-epithelium of the OCT images (presented in Table 1 and described in details in supplementary Table 2 and 3). The primary features were selected for image based disease classification following Ughi et al.[14]

Disease classification by statistical analysis of OCT image features: The NOM and disease classes based on the features were then classified using 'Classification Learner' app of MATLAB 2015a version. Students' two tailed t-test with 95% confidence interval was also performed between each of the disease classes and the NOM to identify disease specific important features separately. Firstly principal component analysis (PCA) followed by linear discriminant analysis (LDA) using 20 principal components were performed utilizing OCT intensity and textural feature towards global classification of oral tissues. Since there were significant overlapping in oral PMDs and normal condition (Fig. 1), SVM classification followed by sequential feature selection technique using mRMR for feature subset selection was also performed[21]. The insignificant features which were not utilized further have been presented in Table 2, while the significant ones were proposed as QIB. Linear, quadratic and cubic kernels variants of SVM were used during classification following 10 fold cross-validation.

OCT-Histology image correlation: Structural correlation between the in vivo OCT images and corresponding H&E image were also performed (Fig. 1). Inter-rater agreement to assess the strength of agreement for disease class identification from histology images were shown using Kappa scoring and presented, as detected by two expert onco-pathologist in blind manner (Table 3).

Table 1. Features considered during OCT image analysis

Feature Type	Features Considered	Reference
Intensity	1. Mean Gray 2. Median Gray 3. Standard Deviation Gray 4. Entropy Gray 5. Coefficient of variance Gray 6. Skewness Gray 7. Kurtosis Gray 8. Variance Gray	26
Texture	9.Contrast of Gray level co-occurrence matrix (GLCM) 10. Correlation of GLCM 11. Energy of GLCM 12. Entropy of GLCM 13. Homogeneity of GLCM 14. Cluster shade 15. Cluster prominence 16.Information measures of correlation 17. Max Probability 18. Sum of entropy 19. Sum of variance 20. Difference entropy 21. Local binary pattern (LBP), mean 22. LBP, standard deviation	14, 26

Table 2. Irrelevant features identified for oral lesion classification after mRMR sequential feature selection technique during oral lesion classification

Diseases Classified	Feature not used (Feature Number used in Table 1)
NOM vs. OSF	2, 3, 4, 8, 14,
NOM vs. OLK	14
NOM vs. OSCC	0
OLK vs. OSCC	10, 13
OSF vs. OSCC	8, 19, 20
OLK vs. OSF	1, 8, 19, 20
NOM vs. OSF vs. OSCC	1, 4, 8, 18,19, 20
NOM vs. OLK vs. OSCC	8, 9, 20
NOM vs. OLK vs. OSF vs. OSCC	4,9

Table 3: Strength of agreement for disease identification from histology images to address inter-observer variability

Disease	Kappa Score (κ)	standard error of kappa (SEκ)	95% Confidence Interval	The strength of Agreement
NOM	0.857	0.136	0.59-1.00	Very Good
OLK	1.000	0.000	1.00-1.00	Perfect
OSF	0.941	0.058	0.83-1.00	Very Good
OSCC	1.000	0.000	1.00-1.00	Perfect

FTIR data acquisition: The study was performed using Nicolet 6700 spectrometer (Thermo Fisher, USA). Spectral data was acquired in transmission mode using acetone treated dried deparaffinized unstained sections. The tissues were dried using 5 minute of acetone treatment, removed from slides and made into powder form. KBr (Potassium Bromide) pellet of the dried tissues were prepared using 0.02 mg of sample and 2 mg of KBr [27]. One spectrum per sample was taken for each tissue section in KBr pellet. Minimization of tissue specific spectral variation was achieved using mean spectra of three tissue sections for each sample. Hence for 57 surgical samples, 171 spectra were taken. All the FTIR spectra were obtained for the range of 400–4000 cm^{-1} at a resolution of 4 cm^{-1} with 32 scans. An 8-mm aperture diameter and DTGS detector was used during data acquisition.

Statistical analysis of FTIR data: During spectral pre-processing, primary feature selection was performed for the spectral band between 'fingerprint' region, 1800 – 900 cm^{-1}.[6] Then 1st Savitzki-Golay differentiation was applied for spectral smoothing (in first order differentiated spectra, differentiation order was 1, polynomial order was 2 and number of filter coefficients was 9). Further, spectrum-wise vector normalization as well as variable-wise maximum normalization was performed. Then PCA was performed using 50 principal components followed by LDA. Tissue hydration status even after tissue processing and complete drying of the tissue was evaluated by feature selection of the spectra between (a) 1600-1800 cm^{-1}, (b) 2400-2000 cm^{-1} and (c) 3700 – 3000 cm^{-1} after rubberband like baseline correction (RBBC). The analyses were executed using 'IRootLab'[28], MATLAB toolbox used for vibrational spectroscopy in MATLAB R2015a (MathWorks, USA). Second derivative of average spectra of each condition was plotted using the OMNIC 9 software.

Result and Discussion

Biopsy based histopathological classification for oral pre-cancers needs value addition due to lack of precise disease specific marker[19] and high inter-observer variability[19, 29]. In this regard multimodal information generation and their logical corroboration may be effective as performed in this study by using OCT imaging along with relevant HC, IHC and FTIR studies.

Present study utilized OCT imaging for specifying structural information, which was then correlated with histological findings (Fig 1). In Fig 1a and Fig1e architectural features like distinct keratin layers above supra-basal layer and prominent grooves of rete pegs were evident in NOM condition, while in Fig 1b and Fig 1f increase in keratin thickness and hyperplasia were clearly visible, which were characteristic histological features of OLK. In OSF cases (Fig 1c and Fig 1g), atrophic rete-pegs and increased collagen deposition was noticed in sub-epithelial region. In cases of OSCC (Fig 1d and Fig 1h), increased numbers of blood vessels were clearly seen, which might be due to neo-angiogenesis.

Although this study correlated OCT with histology, but still expert based interpretation was needed for diagnosis. The intensity and textural features of OCT were therefore analyzed towards value addition to the process of oral pre-cancer and cancer differential diagnosis as well as to boost computer aided diagnostic (CAD) technique. The texture and intensity features based classification of OCT images (Table 1) using PCA-LDA score plots (Fig 2), depicted significant overlapping in between NOM and PMD conditions. Therefore, SVM based two class disease classification was further performed. The results suggested that selected OCT features could differentiate the lesions with high sensitivity and specificity, mostly with 90% overall accuracy. Table 4 presents the classification performance of SVM with different kernels, with 10 fold cross validation. Cubic and quadratic kernels were found to be more efficient than linear kernel during classification. All the lesions could be classified using quadratic SVM with 82.6%

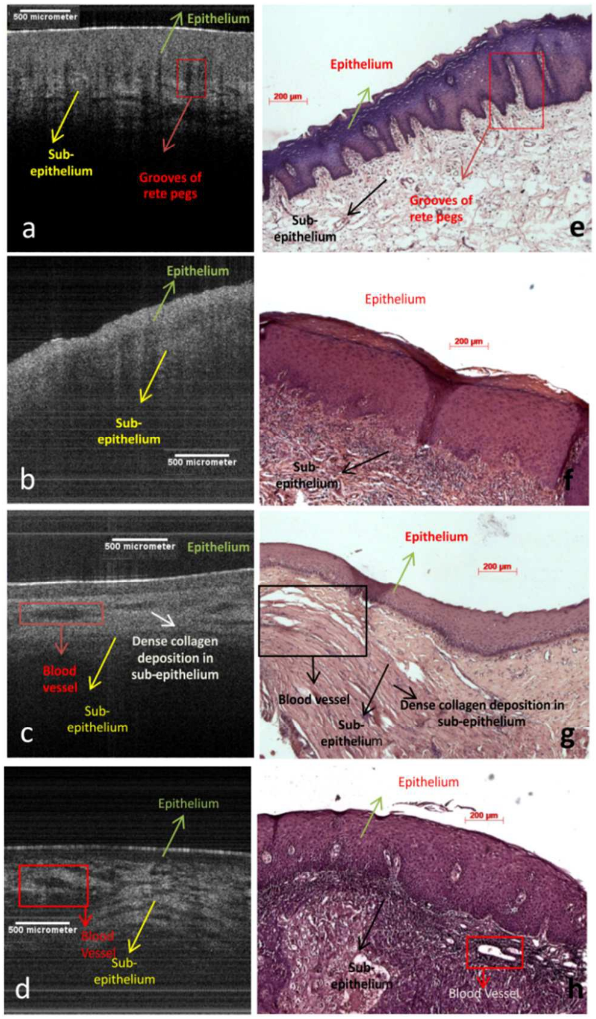

Fig 1. In vivo OCT images of (a) NOM (b) OLK, (c) OSF (d) OSCC and corresponding H&E images (at 5x magnification) (e) NOM) (f) OLK (g) OSF and (h) OSCC depicting structural correlation

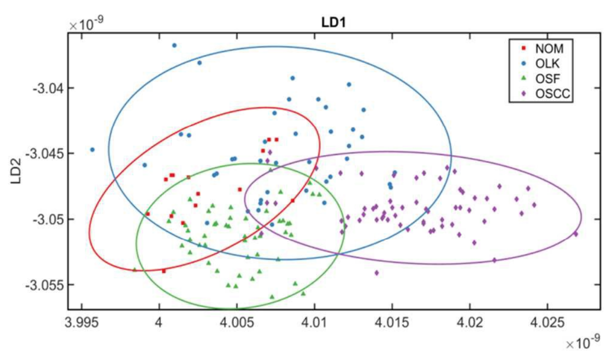

Fig 2. LDA score plot of OCT intensity and textural feature using 20 principle components after PCA-LDA with confidence ellipse representing confidence interval at 80%

accuracy after optimization of classifiers, when four class classifications was performed. The confusion matrix has been presented in Fig 4. Further, sequential feature reduction based attribute selection was then performed towards optimal classification of OCT features. Specific feature subsets, not selected during two class disease classification (provided in Table 2), were not used further, and therefore the rest (not mentioned in Table 1) were considered as optimum selected features from OCT images.

When two tailed 't' test with 95% confidence interval was performed between the disease conditions with the OCT features, the information measure was significant to differentiate NOM vs. OSCC and NOM vs. OSF. Skewness of gray value was important for NOM vs. OSCC and NOM vs. OLK. Other statistically significant parameters ($p<0.05$) to delineate NOM and OSCC were mean and median of gray values, entropy of GLCM, cluster shade, cluster prominence and sum of variance. Correlation and homogeneity of GLCM, difference entropy as well as mean and standard deviation of LBP were found to be important to distinguish NOM and OSF. Lower entropy, mean of gray value in OSCC indicated increased homogeneity in OCT images, while low

Table 4: Classification performance of variants of SVM based on intensity and texture features extracted from OCT images

Classification conditions	Classifier Used	Sensitivity (%)	Specificity (%)	Accuracy (%)
NOM vs. OLK	Linear SVM	37.5	85.4	71.9
	Quadratic SVM	81.3	92.7	89.5
	Cubic SVM	**87.5**	**97.6**	**94.7**
NOM vs. OSF	Linear SVM	56.3	98	88
	Quadratic SVM	50	98	86.6
	Cubic SVM	**75**	**98**	**92.5**
NOM vs. OSCC	Linear SVM	81.3	93.8	91.3
	Quadratic SVM	68.8	95.3	90
	Cubic SVM	**81.3**	**96.6**	**93.8**
OLK vs. OSCC	Linear SVM	58.5	85.9	75.2
	Quadratic SVM	80.5	87.5	84.8
	Cubic SVM	**80.5**	**92.2**	**87.6**
OSF vs. OSCC	Linear SVM	90.2	92.2	91.3
	Quadratic SVM	**88.2.2**	**95.3**	**92.2**
	Cubic SVM	88.2	92.2	90.4
OLK vs. OSF	Linear SVM	78	80.4	79.3
	Quadratic SVM	**92.7**	**86.3**	**90.2**
	Cubic SVM	87.8	86.3	87

cluster prominence indicated small variation in gray scale too, as validated from H&E images (Fig 2)[30].

Since biochemical alteration could only be validated by multimodal characterization of oral lesions and selected feature subset obtained from OCT images can only be rechristened to QIBs if significant alterations are present in disease conditions[17], HC and IHC studies were performed and logically corroborated with OCT images towards better disease characterization. The HC and IHC (Fig 3) were effective to elucidate specific local molecular signatures, corroborative with structural information from OCT.

As per HC and IHC observation, increased PAS positivity was obtained (Fig 3b) in OLK than NOM (Fig 3a) and it was in synergy with a previous result, since hyperkeratosis was a signature in OLK.[23] This result is also reflected in OCT (Fig 1b) and histology (Fig 1f). Further the observation on expression of keratin producing cells, sought by CK 10 expression, a marker of early terminal differentiation-cum-maturation was indicative for differential diagnosis.[31] Result showed moderate expression of CK10 in NOM (Fig 3e) and OSF (Fig 3g) while increased expression of this molecule throughout epithelium in OLK (Fig 3f) indicated the presence of immature keratin producing epithelial cells. However, CK10 expression was not evident in OSCC (Fig 3h).

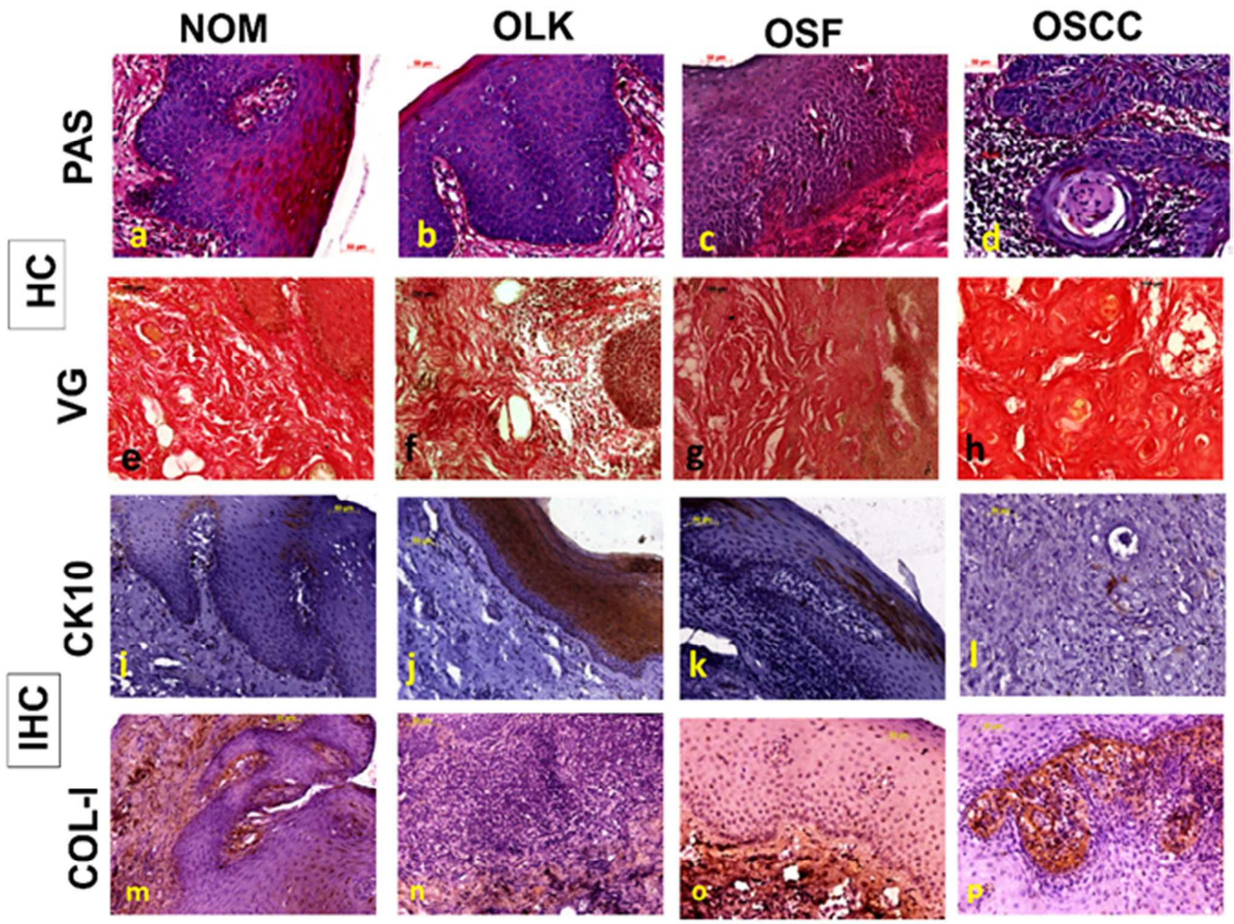

Fig 3. Representative images of each study conditions of PAS stained tissue section of (a) NOM) (b) OLK (c) OSF and (d) OSCC; CK-10 stained tissue section of (e) NOM) (f) OLK (g) OSF and (h) OSCC; VG stained tissue section of (i) NOM) (j) OLK (k) OSF and (l) OSCC; Collagen I stained tissue section of (m) NOM) (n) OLK (o) OSF and (p) OSCC; Microscopic images provided at 20X magnification

Please do not adjust margins

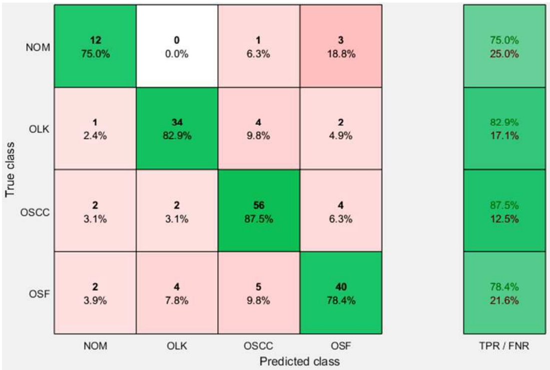

Fig 4. Confusion matrix of multiclass oral lesion classification using intensity and textural features extracted from OCT images by quadratic SVM at 10 fold cross validation (TPR – True Positive Ratio, FNR – False Negative Ratio)

When VG stained sections of OSF were compared to NOM and other conditions, significant increase in collagen deposition was found in lamina propria (Fig 3i-l), as literature suggests that in OSF muscle fibres are replaced by collagen.[32]

Since the epithelium of all OLK cases were found to be immunopositive for CK10, in synergy with the previous studies,[33] it can be deduced that increased lucidity of OLK in OCT image (Fig 1b) was perhaps due to increase in epithelial keratinized cells (Fig 3j) as well as more increased nuclei size than NOM (Fig 3a-b).[8] Again in OSF, distinct lucidity of sub-epithelium (Fig 1c) could be due to increased COL-I expression (Fig 3o). In OSCC, OCT image (Fig 1d) was homogeneous in nature, as distinctness between epithelium and sub-epithelium was minimal. Same observation was also supported by H&E staining (Fig 1h).

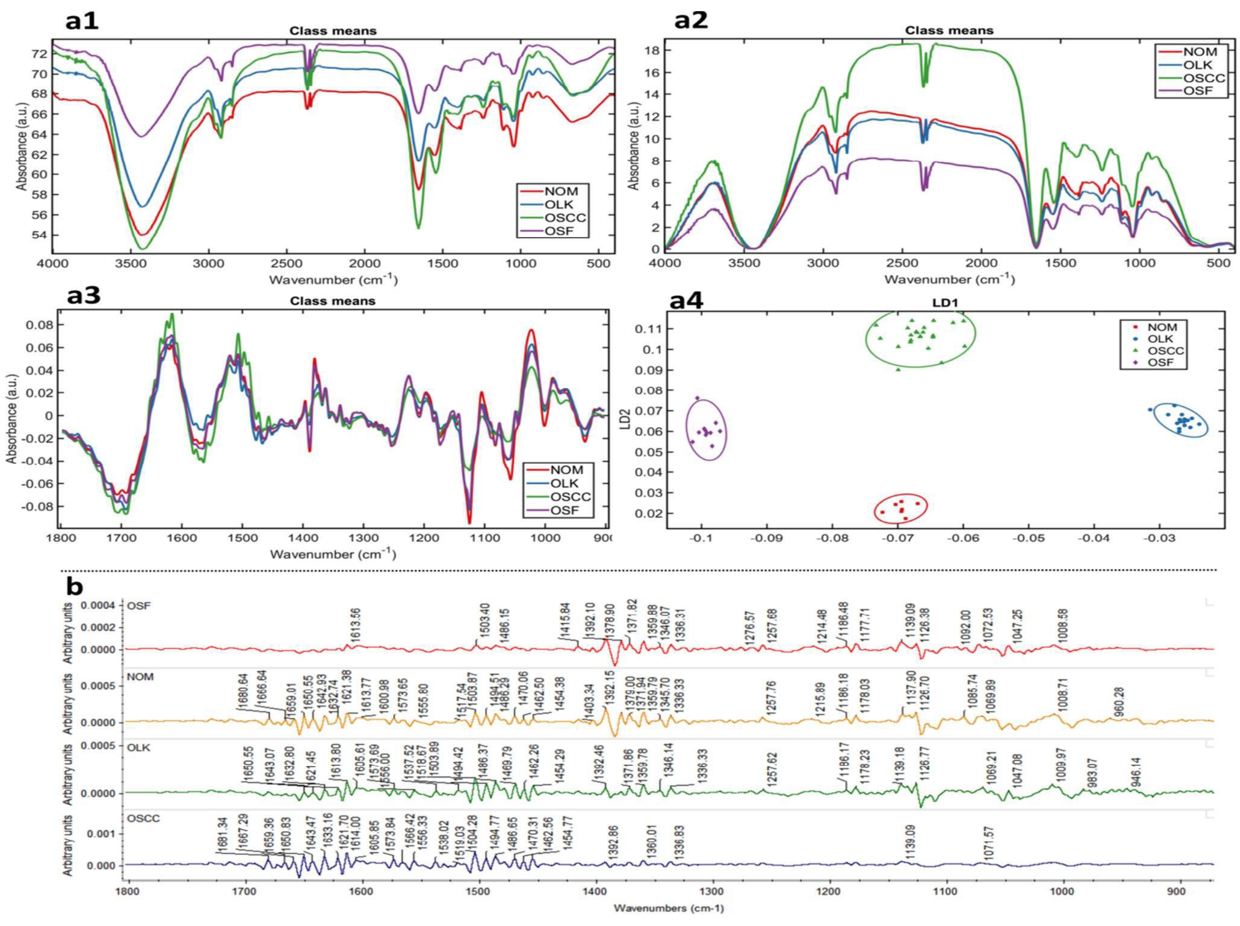

Fig 5. (a1) Mean FTIR spectra of whole region (400-4000^{-1} cm) (a2) Mean spectra of whole region (400-4000^{-1} cm) after rubberband like base like correction (RBBC) (a3) Mean spectra of fingerprint region after RBBC, maximum vector normalization followed by Savitzki-Golay differentiation of 1st Derivative spectra of NOM, OLK, OSF and OSCC (a4) LDA scores plot of pre-processed spectra after mean centering and PCA-LDA with confidence ellipse representing confidence interval at 95% (a.u – arbitrary unit),(b) Second derivative of average FTIR spectra of NOM, OLK, OSF and OSCC

RSC Advances

ARTICLE

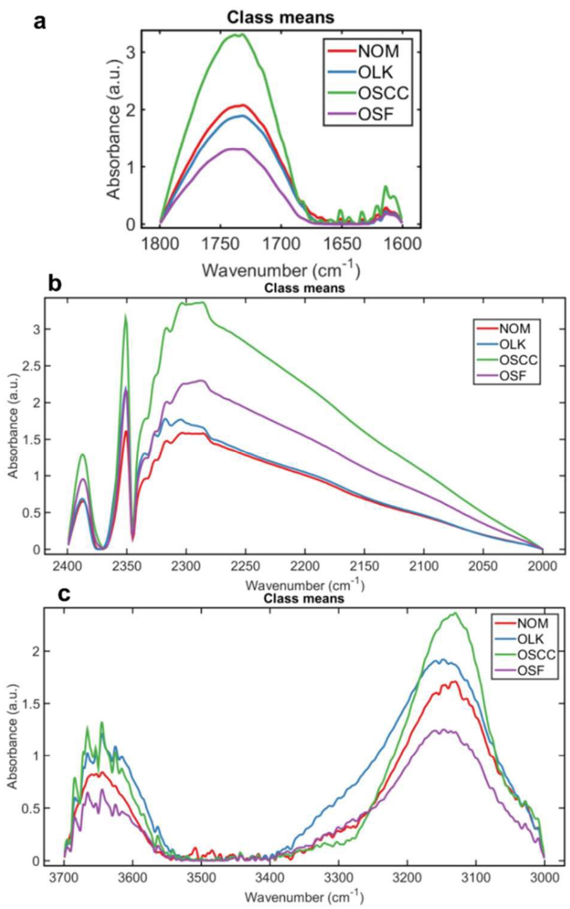

Fig 6. Mean spectra after RBBC of the area between (a) 1600-1800 cm^{-1} (b) 2400-2000 cm^{-1} and (c) 3700 – 3000 cm^{-1} for depicting tissue hydration status of NOM, OLK, OSF and OSCC

As HC and IHC could provide only local information related to some specific molecules of epithelium and sub-epithelium, in addition to these, FTIR (Fig 5 and 6) was performed on these oral pathosis to check whether discriminating signature could be noted for the presence of unique disease specific global chemical alteration. The lesions were therefore tried to be segregated on the basis of global chemical signatures obtained in 'fingerprint' region of FTIR after spectral pre-processing by optimized PCA-LDA. Result presented in Fig 5a suggested that the lesions could be completely segregated when LDA scores plot with confidence ellipse representing confidence interval at 95% were plotted. It suggested significant variations in chemical composition between oral lesions, and thus the notion of disease specific chemical signature was validated. Since a recent review suggested degradation of collagen cores in OSF[34], the cause of disappearance of peaks in OSF area between 1400-1700 cm^{-1} (Fig. 5b) was found possibly due to decrease in skeletal muscle phospholipid and proteins. Although collagen fibres are rich in proline and/or hydroxyproline[35], amount of these amino acids along with glycine was found to be decreased in OSF[36]. The peak picking from second derivative spectra of the same region '1800 - 900 cm^{-1}' for understanding minute chemical changes in each disease condition, Fig 5 depicted only minute alteration existed between NOM and OLK, but it was noted that these two could be classified from OCT images with highest accuracy (Table 4).

This result may be attributed to alteration in tissue hydration status that affects the scattering in OCT, as evident from Fig 6. When mean spectra after RBBC of areas between 1600-1800 cm^{-1}, 2000-2400 cm^{-1} and 3700–3000 cm^{-1} were considered, it was observed that OSCC possess higher content of bound water than normal condition, which was also in synergy with a recent study[7]. It was also evident from Fig 5 that bound water content in OLK was lesser than OSCC, but higher than NOM. Among all these oral lesions, bound water content was found to be least in OSF, which was not found to be reported in any previous studies. It was evident from Fig 5b that, in OSCC many peaks were found to be dissolved in the area of carbohydrates including glycogen (1200 – 900 cm^{-1}). This finding was in synergy with a previous study[6] and also validated with PAS positivity of the tissue sections (Fig 3 a-d). Depletion of glycogen and associated proteins were found to be the major chemical attributes of OSCC, which also might be the underlying cause of homogeneity in pre-processed FTIR peak in the area of 900-1500 cm^{-1} (Fig 4b).

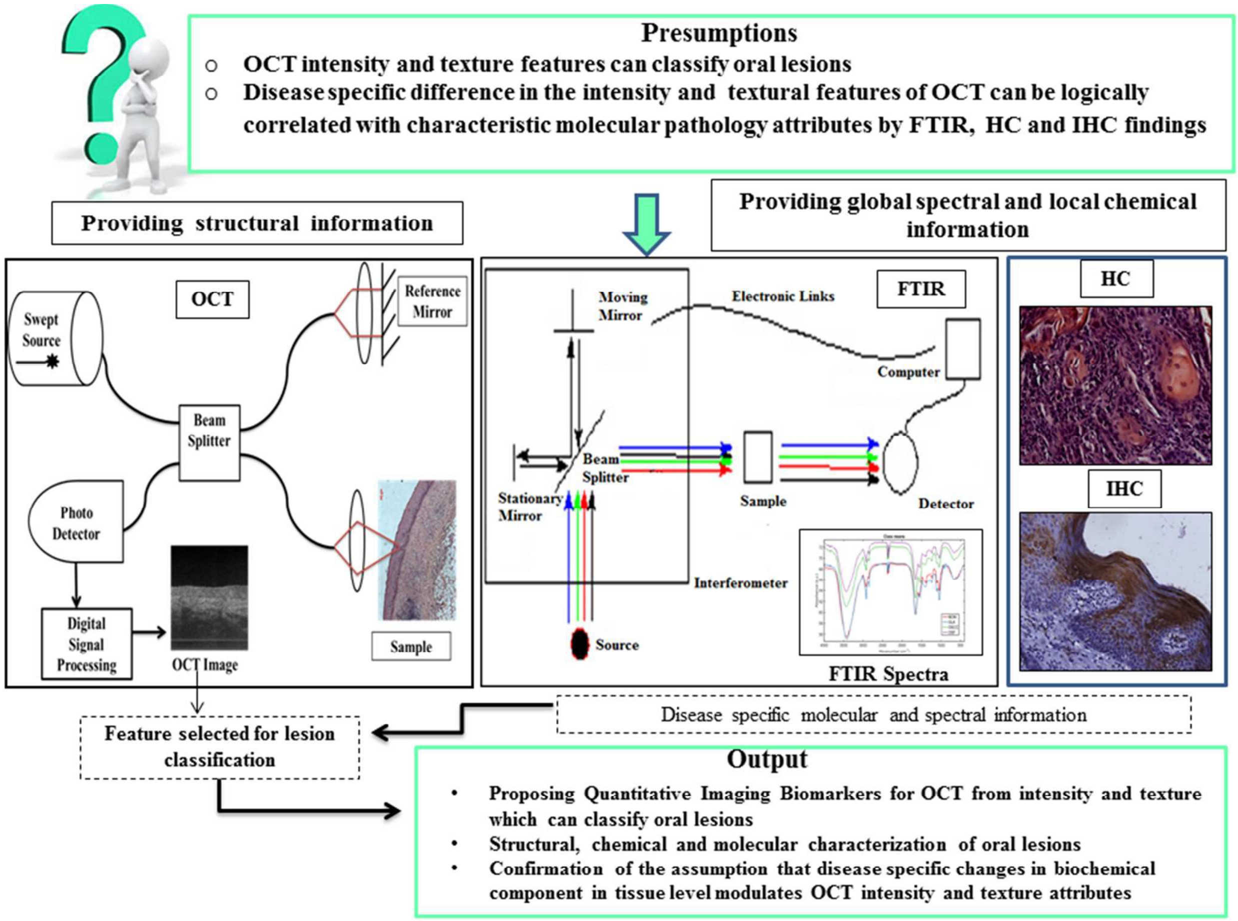

Fig 7. Representing multimodal characterization of oral lesions with plausible informational convergence endorsing complementarity of methods

Since biochemical alteration could be validated by multimodal characterization of oral lesions,[17] the optimum selected features from OCT images therefore can be finally proposed as QIBs. Results thus also helped to prove both the proposed hypotheses that, oral lesions can be subjectively distinguished and characterized from the multimodal information obtained from OCT- histology-HC-IHC and objectively classified with the aid of intensity and texture features of OCT images. It can also be substantiated that, difference in the disease specific epithelial and sub-epithelial intensity and texture were due to chemical alteration of epithelium and sub-epithelium in different lesions. Hence it may be concluded that OCT information can be logically corroborated with FTIR, HC and IHC. The presumptions used to devise the study, the approach for addressing the research questions and a crisp outcome from meaningful integration of quantitative as well as qualitative knowledge obtained in this study has been depicted in Fig 7.

Conclusion:

This study convolved multimodal evaluation and classification of oral lesions, through integration of complementary information obtained from non-invasive techniques like OCT and invasive techniques like FTIR, HC and IHC of biopsied tissues. OCT provided morphological as well as optical features, FTIR documented global chemical signatures while HC and IHC showed local expression of specific biochemical components in epithelial and sub-epithelial compartment. This proof-of-concept study also showed efficacy of intensity and textural features extracted from OCT images towards optimal painless diagnostic segregation of oral diseases with high sensitivity and specificity, which may mitigate challenges associated with inter - observer variability in histopathological interpretation. Lesion specific biochemical changes highlighted few alterations in epithelial and sub-epithelial characteristics, on the basis of which OCT attributes could be re-christened to QIBs for oral lesion differentiation. Keratin associated epithelial and collagen associated sub-epithelial changes were found to be most significant in oral lesion pathogenesis through this multimodal tissue characterization study. This study therefore can be considered as a new hope to understand and differentiate oral lesions from multimodal imaging and analysis as well as system pathology approach[37].

Acknowledgement:

The authors would like to acknowledge financial support of MHRD, Government of India, New Delhi from IIT, Kharagpur (IIT/SRIC/SMST/IAN/2013-14/222). The authors would also like to thank Mr. Debnath Das and Mr. B Mohan Rao for their help during tissue processing and FTIR data acquisition respectively.

References:

1. W. Jerjes, T. Upile, B. Conn, Z. Hamdoon, C. S. Betz, G. McKenzie, H. Radhi, M. Vourvachis, M. El Maaytah, A. Sandison, A. Jay and C. Hopper, *The British journal of oral & maxillofacial surgery*, 2010, **48**, 18-25.
2. J. Gallwas, L. Turk, K. Friese and C. Dannecker, *Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 2010, **36**, 624-629.
3. A. Barui, P. Banerjee, R. Patra, R. K. Das, S. Dhara, P. K. Dutta and J. Chatterjee, *Journal of biomedical optics*, 2011, **16**, 026010.
4. E. Regar, A. van Leeuwen and P. W. Serruys, *Optical coherence tomography in cardiovascular research*, CRC Press, 2007.
5. N. Brill, J. Riedel, B. Rath, M. Tingart, H. Jahr, M. Betsch, V. Quack, T. Pufe, R. Schmitt and S. Nebelung, *Biomedical optics express*, 2015, **6**, 2398-2411.
6. S. Banerjee, M. Pal, J. Chakrabarty, C. Petibois, R. R. Paul, A. Giri and J. Chatterjee, *Anal Bioanal Chem*, 2015, **407**, 7935-7943.
7. E. M. Barroso, R. W. Smits, T. C. Bakker Schut, I. ten Hove, J. A. Hardillo, E. B. Wolvius, R. J. Baatenburg de Jong, S. Koljenovic and G. J. Puppels, *Analytical chemistry*, 2015, **87**, 2419-2426.
8. J. J. Rosette, M. J. Manyak, M. G. Harisinghani and H. Wijkstra, *Imaging in Oncological Urology*, Springer Science & Business Media, 2008.
9. D. Sheet, A. Chaudhary, S. P. K. Karri, D. Das, A. Katouzian, P. Banerjee, N. Navab, J. Chatterjee and A. K. Ray, *Journal of biomedical optics*, 2013, **18**, 090503-090503.
10. Y. Gan, W. Yao, K. M. Myers, J. Y. Vink, R. J. Wapner and C. P. Hendon, *Biomedical optics express*, 2015, **6**, 1090-1108.
11. D. Sheet, S. Banerjee, S. P. K. Karri, S. Bag, A. Anura, A. Giri, R. R. Paul, M. Pal, B. C. Sarkar and R. Ghosh, 2014. DOI:10.1109/ISBI.2014.6868137
12. Z. Hamdoon, W. Jerjes, G. McKenzie, A. Jay and C. Hopper, *Photodiagnosis and photodynamic therapy*, 2015, DOI: 10.1016/j.pdpdt.2015.07.170.
13. Z. Hamdoon, W. Jerjes, T. Upile, G. McKenzie, A. Jay and C. Hopper, *Photodiagnosis and photodynamic therapy*, 2013, **10**, 17-27.
14. G. J. Ughi, T. Adriaenssens, P. Sinnaeve, W. Desmet and J. D'Hooge, *Biomedical optics express*, 2013, **4**, 1014-1030.
15. P. Pande, S. Shrestha, J. Park, M. J. Serafino, I. Gimenez-Conti, J. Brandon, Y. S. Cheng, B. E. Applegate and J. A. Jo, *Journal of biomedical optics*, 2014, **19**, 086022.
16. J. A. Kiernan, *Histological and histochemical methods*, Med Council on Alcohol, 1990.
17. J. W. Prescott, *Journal of digital imaging*, 2013, **26**, 97-108.
18. P. Koturwar, S. Girase and D. Mukhopadhyay, *arXiv preprint arXiv:1503.07477*, 2015.
19. S. Banerjee and J. Chatterjee, *Springer Science Reviews*, 2015, **3**, 127-136.
20. N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*, Cambridge university press, 2000.
21. H. Peng, F. Long and C. Ding, *IEEE transactions on pattern analysis and machine intelligence*, 2005, **27**, 1226-1238.
22. H. Krishna, S. K. Majumder, P. Chaturvedi, M. Sidramesh and P. K. Gupta, *Journal of biophotonics*, 2014, **7**, 690-702.
23. B. W. Neville and T. A. Day, *CA: a cancer journal for clinicians*, 2002, **52**, 195-215.
24. K. A. Moutasim, V. Jenei, K. Sapienza, D. Marsh, P. H. Weinreb, S. M. Violette, M. P. Lewis, J. F. Marshall, F. Fortune, W. M. Tilakaratne, I. R. Hart and G. J. Thomas, *The Journal of pathology*, 2011, **223**, 366-377.
25. J. J. Pindborg, P. Reichart, C. Smith and I. Van der Waal, *Definitions and explanatory notes*, Springer, 1997.
26. M. Sonka, V. Hlavac and R. Boyle, *Image processing, analysis, and machine vision*, Cengage Learning, 2014.
27. A. H. Colagar, M. J. Chaichi and T. Khadjvand, *Journal of biosciences*, 2011, **36**, 669-677.
28. J. Trevisan, P. P. Angelov, A. D. Scott, P. L. Carmichael and F. L. Martin, *Bioinformatics (Oxford, England)*, 2013, **29**, 1095-1097.
29. O. Kujan, A. Khattab, R. J. Oliver, S. A. Roberts, N. Thakker and P. Sloan, *Oral oncology*, 2007, **43**, 224-231.

30. X. Yang, S. Tridandapani, J. J. Beitler, D. S. Yu, E. J. Yoshida, W. J. Curran and T. Liu, *Medical physics*, 2012, **39**, 5732-5739.

31. S. Bag, M. Pal, A. Chaudhary, R. K. Das, R. R. Paul, S. Sengupta and J. Chatterjee, *Journal of clinical pathology*, 2015, **68**, 605-613.

32. T. Rooban, T. R. Saraswathi, F. H. Al Zainab, U. Devi, J. Eligabeth and K. Ranganathan, *Indian journal of dental research*, 2005, **16**, 131-134.

33. T. Fillies, M. Jogschies, J. Kleinheinz, B. Brandt, U. Joos and H. Buerger, *Oncology reports*, 2007, **18**, 639-643.

34. V. Kamath, *Saudi Journal of Oral Sciences*, 2014, **1**, 57.

35. C. Vidal Bde and M. L. Mello, *Micron (Oxford, England : 1993)*, 2011, **42**, 283-289.

36. I. Y. Huang and T. Y. Shieh, *Gaoxiong yi xue ke xue za zhi = The Kaohsiung journal of medical sciences*, 1989, **5**, 162-171.

37. J. Costa, *Molecular oncology*, 2012, **6**, 27-32.