A Multi-Scale Context Aggregation Enriched MLP-Mixer Model for Oral Cancer Screening from Oral Sub-epithelial Connective Tissues

Sawon Pratiher[§], Subhankar Chattoraj[†], Debaleena Nawn[§], Mousumi Pal[ς]
Ranjan Rashmi Paul[ς], Hubert Konik[‡], and Jyotirmoy Chatterjee[§]
[§]*Indian Institute of Technology Kharagpur, India.*
[†]*Université Jean Monnet, Saint-Étienne, Univ. Lyon, France.*
[ς]*Guru Nanak Institute of Dental Sciences and Research Kolkata, India.*
[‡]*Univ. Lyon, UJM-Saint-Etienne, CNRS, Institut d' Optique Graduate School,*
Laboratoire Hubert Curien UMR 5516, F-42023, Saint-Etienne, France.

***Abstract*—This treatise proposes a dilated convolutional multi-layer perceptron (MLP)-mixer model (hereafter referred to as DiCoMLP-Mixer) for oral connective tissue (OCT) grading. The proposed DiCoMLP-Mixer framework comprehends dense multi-scale contextual feature representation by enabling exponential receptive field expansion without sacrificing resolution via dilated convolutions. The MLP-mixer backbone in the DiCoMLP-Mixer architecture leverages the attention mechanism of a transformer model for saliency abstraction in oral mucosa histopathological images (OMHIs) by the spatial encoding of OMHI patches. This study focuses on the oral mucosa's sub-epithelium region, which has considerable clinical significance but is understudied in the literature. DiCoMLP-Mixer's exhaustive experimental validation on two OCT layers of the sub-epithelium, namely papillary (L_1) and reticular (L_2) for oral cancer (OC) grading, classifying the three major oral potentially malignant disorders (OPMDs) and OC from healthy OCTs. The ablation study with the existing MLP-mixer model evinces enhanced OC screening performance, while Grad-CAM heatmaps exhibit DiCoMLP-Mixer's consistent clinical saliency for precise OC detection.**

***Index Terms*—Bright-field microscopy; oral cancer; OPMDs; sub-epithelium; MLP mixer; dilated convolution; explainable AI.**

I. Introduction

GLOBOCAN reports an estimated 377,713 new OC cases and 177,757 OC-related deaths globally in 2020 [1], making OC one of the most frequent malignancies worldwide [2]. Similarly, Oral Cancer Foundation reports an estimated 53,000 new OC cases registered in the United States in 2019 [1]. Excessive alcohol consumption and widespread use of betel quid and tobacco in South Asia's low- and middle-income countries account for two-thirds of the global OC incidence [3]. OC is identified by late diagnosis, high fatality, morbidity, low survival rates, and costly treatment, especially in later stages. Further, the lack of public awareness and health professionals' understanding of OC markers is responsible for its late identification [4].

OC is clinically diagnosed by distinguishing between malignant and ulcerated lesions in the oral cavity. During prognosis, the fundamental difficulty for clinicians is detecting benign lesions that mimic malignant ones. Pre-existing OPMDs like oral submucous fibrosis (OSF), oral leukoplakia (OLKP), and erythroplakia are the most common OC precursors [5]. Early-stage OPMD diagnosis reduces OC's risk and tumor progression by arranging timely clinical treatment regimens for patients' survival rate improvement [6]. Several image modalities such as microscopic [6], [7], [8], hyperspectral [9], computed tomography [10], autofluorescence [11], fluorescence [12], and standard white light image of oral cavity structures [13] has been employed for OC screening. The gold standard for detecting OPMDs and aberrant cellular activity is the light-field microscopic histopathological assessment of biopsy samples [8]. However, manual diagnosis of OPMD lesions is tedious, subjective, and dependent on the clinician's expertise [7]. Computer-aided diagnostic (CAD) systems with high sensitivity are imperative to overcome the hurdles mentioned above during OC detection and reduce the risk of early-stage malignant lesions being wrongly categorized as benign.

The prior works on OC diagnosis from OMHIs comprise both shallow feature-based machine learning (ML) and deep learning (DL) methods. Hand-crafted features include statistical eccentricity and compactness [14], morphological [15], textural feature analysis [16], multifractal alterations [17], histogram [18], gray-level co-occurrence matrix (GLCM) [19], local binary pattern (LBP) [20], higher-order statistics (HOS) [6], Gabor wavelet [21], and Fuzzy statistics [22], with ML classifiers like support vector machine (SVM) with radial basis function (RBF) kernel, Bayesian classifier, k-means, Gaussian mixture model (GMM) and Fuzzy c-means clustering (FCM) [22]. DL methods include convolutional neural networks (CNN) [23]. The literature on OPMD analysis is abundant on morphological alterations in the epithelial region, and to the best of our knowledge, sub-epithelial region analysis has received little attention. Further, the prior art in [24], [25] highlights that collagen fibers of the Extracellular Matrix in the subepithelial layer undergo structural change with the onset of malignancy, with substantial microscopic variations in the sub-epithelial collagen fibers, have been observed for OSF groups [26]. Quantitative analysis of collagen evolution is essential to assess OSF progression into OSF with dysplasia (OSFD), and OSF without dysplasia [27].

The sparsely explored sub-epithelial oral mucosa (having high clinical importance) and the need for robust CAD to characterize the collagen adaptations from healthy (NOM) tissues to OPMDs (OSF, OSFD, and OLKP) and OC classes motivate us to investigate the recent state-of-the-art DL models for oral cavity tissue characterization. Oral oncopathologists guide the cropping of tissue-index transmission patches (TITPs) from OMHIs belonging to the OCT layers of the sub-epithelium at two different zones, namely papillary (L_1) and reticular (L_2). We proposed a multi-scale context aggregation enriched MLP-mixer model (hereafter referred to as DiCoMLP-Mixer) on

ISBN: 978-1-6654-6798-8

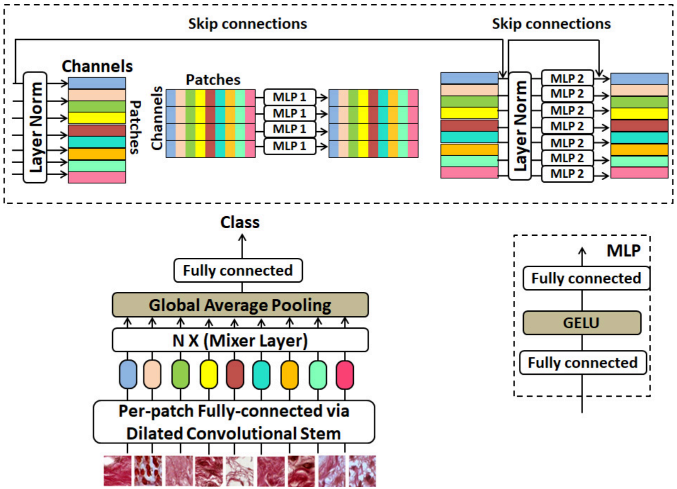

Figure 1: OC screening pipeline showing data acquisition, pre-processing, extraction of OMHIs and TITPs, and schematic diagram of the DiCoMLP-Mixer framework.

these TITPs for OMHI grading. The DiCoMLP-Mixer model uses dilated convolution-based patchify stem to comprehend the multi-scale spatial-deformation dynamics.

The OC screening workflow containing the modular level schematic of the DiCoMLP-Mixer framework is shown in Fig. 1. The following research contributions are derived from DiCoMLP-Mixer's design, application, and assessment.

- DL-based OC screening using OMHIs collected from sub-epithelial OCTs (clinical value for oral onco-pathologists) is sparsely explored in the literature.
- Extensive inter-class classification of five study groups: healthy, OPMDs (OSF, OSFD, OLKP), and OC are analyzed for OMHIs from two OCT (L_1 and L_2) layers of the sub-epithelium.
- Proposed DiCoMLP-Mixer model incorporates a dilated convolutional patchify stem on the MLP-Mixer [28].
- Grad-CAM-based saliency heatmaps evince rich representation learning and provide locally interpretable underlying insights into DiCoMLP-Mixer's predictions.

The remaining paper is structured as follows: Section II describes the dataset and DiCoMLP-mixer model. Section III discusses the experimental results. Section IV concludes the paper.

II. Material and Methods

A. Dataset Description

*1) **Study Design and Data Acquisition**:* The experimental study **(GNIDSR/IEC/07/16)**, was approved by the institutional ethics committee. All data collection was done in full conformity with the Indian Medical Association's ethical principles and standards, including the World Medical Association's and the Helsinki Declaration's ethical principles and guidelines, and with the informed permission of all patients. Incisional oral biopsy specimens from the buccal mucosa of patients were taken at the Guru Nanak Institute of Dental Science and Research (GNIDSR) in Kolkata, India. Expert onco-pathologists histopathologically graded these specimens following Haematoxylin and Eosin staining, and co-morbid samples were excluded. The tissue specimens for the control group (healthy) were taken from the disto-buccal region of the mucoperiosteal flaps. They were raised for trans-alveolar extraction of an impacted mandibular third molar tooth from healthy people with no clinical symptoms of OPMD or oral habits. Our earlier paper [17] detailed the pre-processing of the biopsy specimens for microscopic slide imaging using a bright-field inverted optical microscope (Zeiss Observer.Z1, Carl Zeiss, Germany) and an attached CCD camera (AxioCam MRC, Carl Zeiss) under 40× objective (with pixel granularity: $0.157\mu m$ and with a final magnification of 400×).

*2) **Sub-epithelium Zoning and TITP Generation**:* A loosely defined boundary segregated the sub-epithelial OCT into two layers: papillary (L_1) and reticular (L_2). The papillary layer under the epithelium comprises tiny, loosely packed collagen fibers, and the lower reticular layer is mainly made up of firmly dense collagen fibers [29]. The clinical significance of (L_1) and (L_2) zones in OC progression studies and their extraction followed by generation of TITPs for the current study is explained in detail in our earlier work in [17]. Table I summarize the TITP statistics.

Table I: Dataset statistics.

Class	**No. of. OMHIs**	**No. of TITPs**	
		L_1	L_2
NOM	35	98	98
OSF	36	101	96
OSFD	38	108	101
OLKP	28	99	152
OC	30	92	110
Total	**167**	**498**	**557**

B. Proposed DiCoMLP-Mixer Model

MLP-Mixer [28] has been used as the backbone architecture for proposed DiCoMLP-Mixer. We have integrated a dilated convolutional patchify stem to this MLP-Mixer to encode computationally efficient multi-scale contextual feature abstraction by exponential receptive field expansion. The brief functionality of DiCoMLP-Mixer's different modules is explained below:

C. Backbone MLP-Mixer for Attention Mechanism

The layers in modern MLP-based DL architectures mix features at a specific spatial location, between several spatial locations, or both simultaneously. The per-location (channel-mixing) and cross-location (token-mixing) processes are separated in the MLP-mixer architecture [28]. A sequence (S) of non-overlapping image patches is given as input to an MLP-mixer. Each patch is projected to a specified hidden dimension (C) to produce a two-dimensional real-valued input table, $\mathbf{X} \in \mathbf{R}^{S\times C}$. The number of patches for an input image of size (H, W) and patch size (P, P) is $S = HW/P^2$. With the same projection matrix, all of these patches are linearly projected. The MLP-mixer comprises multiple layers of equal size, each containing two MLP blocks. The first is the token-mixing MLP, which works on $\mathbf{X}$ columns (i.e., it's applied to a transposed input table $X^\top$), maps $\mathbf{R}^S \mapsto \mathbf{R}^S$, and is shared across all columns. The second is the channel-mixing MLP, which acts on rows of $\mathbf{X}$ and maps $\mathbf{R}^C \mapsto \mathbf{R}^C$. It is shared across all rows. Each MLP block has two fully-connected layers, and nonlinearity (such as GELU) is applied to each row of the input data tensor individually. Finally, the MLP-mixer employs a typical classification head with a linear classifier and a global average pooling layer. MLP-mixer details can be found from [28].

D. Dilated Convolutions for Multi-scale Contextual Features

Dilated convolutions enable computationally efficient exponential receptive field expansion for multi-scale contextual feature learning without compromising resolution [30].

III. Experimental Results and Discussion

A. Experimental Setup

The MLP-Mixer [28] and proposed DiCoMLP-Mixer models are implemented in Python using Tensorflow and Keras libraries.The models are trained for 200 epochs using Adam optimizer with batch size = 256, learning rate = 0.005, and weight decay factor = 0.0001. Sparse categorical cross-entropy and accuracy are monitored over 1000 epochs for training convergence. Xaviar uniform initializer is used for model parameter initialization, and early stopping is envisaged to alleviate overfitting. The model inputs are resized to 128x128x3 with patch size = 6x6, embedding dimension = 64, No. of heads in the attention module = 2X embedding dimension, and No. of transformer blocks = 8.

B. Training/Testing Protocol and Performance Evaluation

The TITPs obtained for (L_1) and (L_2) zones are split randomly in the ratio of 80**:**20 for training and validation. All the model predictions output can be categorized as TP (true-positive), TN (true-negative), FP (false-positive), and FN (false-negative). Accordingly, different performance evaluation metrics (PEM) such as accuracy (Acc.), precision (P), recall (R), and F1 score (F1) are computed for method validation. They are defined as: $Acc. = \frac{TP+TN}{TP+TN+FP+FN}$, $P = \frac{TP}{TP+FP}$, $R = \frac{TN}{TN+FN}$, and $F1 = \frac{2TP}{2TP+FP+FN}$. Here, TP/FP (and TN/FN) signify the cardinality of correct/incorrect predictions for the positive (and negative) class, respectively.

C. Classification Results and Ablation Study

Table II summarize the class-specific PEM scores averaged over the 5-folds. The ablation study concerning the introduction of dilated convolutions to the existing MLP-Mixer [28], and DiCoMLP-Mixer's PEM scores outperform the MLP-Mixer on TITP sets of both L_1 and L_2 OCT layers. Moreover, the reticular's (L_2) significance in OC progression analysis seems to be more compared to papillary (L_1).

D. Comparative Evaluation

We performed a comparative analysis of our DiCoMLP-Mixer framework with the prevalent state-of-the-art methods, and contemporary pre-trained DL models, which have achieved great success in various image processing tasks [31].

*1) **Comparison with Pre-trained Models**:* The pre-trained networks are trained on a large-scale database of more than a million images, and 1000 classes have inherently powerful rich features [31]. We have fine-tuned the deeper layers to adapt new features pertinent to the OC screening. Columns 2 in Table III exemplify the model complexity regarding the number of trainable parameters, while columns 3 to 4 tabulate the TITP grading accuracy for L_1 and L_2 zones, respectively, of sub-epithelial OCT. From Table III, DiCoMLP-Mixer exhibit the highest classification accuracy with reduced network depth and comparable trainable parameters and outperform its nearest competitor, DenseNet, by 3.11% and 2.36% on L_1 and L_2 zones, respectively.

*2) **Comparison with the State-of-the-art**:* A brief review of the prevalent OC screening techniques from OMHIs and their performance comparison is summarized in Table IV. To the best of our knowledge, this work is one of the very few research to address a broader study of five classes: healthy, OPMDs (OSF, OSFD, OLKP), and OC and the first to explore MLP-based DL models for OC taxonomy.

Table II: PEM (%) for MLP-Mixer [28] and DiCoMLP-Mixer.

Architecture	**Dataset**	**Class**	**P**	**R**	**F1**	**ACC**
MLP-Mixer [28]	L_1	**NOM**	64.42	88.85	74.78	84.89
		OSF	82.66	64.42	72.47	
		OSFD	89.49	86.47	88.68	
		OLKP	94.05	90.73	93.29	
		OC	90.44	93.28	91.14	
	L_2	**NOM**	89.75	71.67	79.33	87.17
		OSF	67.42	79.33	72.92	
		OSFD	84.23	93.85	89.75	
		OLKP	92.65	96.16	90.77	
		OC	94.57	91.23	95.56	
DiCoMLP-Mixer	L_1	**NOM**	94.08	93.85	93.27	91.29
		OSF	91.11	91.85	91.88	
		OSFD	91.42	94.13	91.33	
		OLKP	92.72	94.42	93.43	
		OC	94.15	94.11	94.96	
	L_2	**NOM**	92.47	92.66	92.68	93.67
		OSF	96.37	94.33	95.62	
		OSFD	94.51	94.7	94.27	
		OLKP	93.65	94.2	94.78	
		OC	94.3	92.9	93.6	

MLP-Mixer [28] run in-house. **BOLD** signifies better performance.

Table III: Accuracy comparison with pre-trained networks.

Architecture	**PM**	L_1	L_2
ResNet-50	25.63	46.87	56.25
ResNet-101	44.70	84.37	89.75
ResNet-152	60.38	84.37	90.62
VGG-16	138.3	87.50	90.17
VGG-19	143.6	81.63	85.93
DenseNet	20.0	88.18	91.31
Inception	23.85	73.43	87.5
Inception-ResNet	55.87	76.56	82.81
Xception	22.91	82.81	87.50
MobileNet	3.53	75.39	78.12
DiCoMLP-Mixer	4.5	**91.29**	**93.67**

PM = number of parameters (in millions).

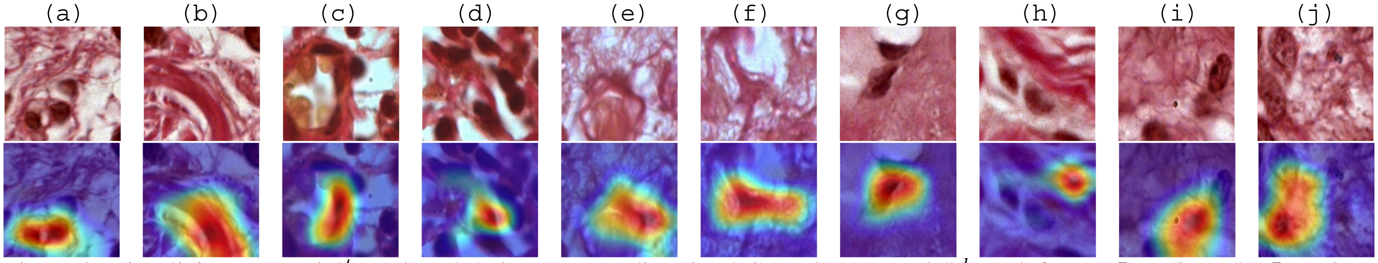

Figure 2: Visualizing TITPs (1^{st} **row)** and their corresponding Grad-CAM heatmaps (2^{nd} **row)** for (a) L_1-NOM, (b) L_2-NOM, (c) L_1-OSF, (d) L_2-OSF, (e) L_1-OSFD, (f) L_2-OSFD, (g) L_1-OLKP, (h) L_2-OLKP, (i) L_1-OC, and (j) L_2-OC. Red indicates area of highest attention and blue indicates area of lowest attention.

Table IV: Prior art comparison.

Ref, Years	**Methodology: (Features + Classifier)**	**NI**	**NC**	**PEMs (%)**		
				AC	SN	SP
[14], 2009	Eccentricity, Compactness, SVM	10	2	88.6	90.4	87.5
[15], 2010	Morphological feature, SVM	32	2	97.1	N.R	N.R
[16], 2010	Textural analysis, SVM-RBF	159	2	88.2	89.9	90.1
	Textural analysis, Bayesian	159	2	90.7	94.7	86
[22], 2012	Fuzzy divergence, morphological feature, bayesian	750	2	96.5	96.4	96.6
	Fuzzy divergence, morphological feature, K-means	750	2	84	84.4	83.2
	Fuzzy divergence, morphological feature, FCM	750	2	89.4	90.1	88.1
	Fuzzy divergence, morphological feature, GMM	750	2	90.3	89.6	91.7
[19], 2018	Histogram, GLCM, SVM	269	2	89.7	94	N.R
[20], 2011	LBP, SVM	158	3	83.5	82.8	87.8
	HOS, SVM	158	3	92.4	94	91.2
[18], 2011	Statistical feature, SVM	30	3	86.6	80.6	97.84
[21], 2011	Gabor Wavelet, SVM	158	3	88.3	N.R	N.R
[6], 2012	HOS, LBP, Fuzzy	158	3	95.7	94.7	98.8
[32], 2011	Statistical and morphological feature, SVM	269	3	93.5	91.6	92.3
[23], 2022	TL, CNN	1200	2	83.81	74.4	89.1
This Work	MLP-Mixe, Dilated Conv. (L_1)	498	**5**	**91.29**	**92.6**	**90.9**
	MLP-Mixer, Dilated Conv. (L_2)	557	**5**	**93.67**	**94.5**	**92.9**

NI: Total No. of images, NC: No. of classes present,
NR: Not reported, TL: Transfer learning.

E. Explainability of DiCoMLP-Mixer's Predictions

Fig. 2 depicts the gradient-weighted class activation mapping (Grad-CAM) explainability of DiCoMLP-Mixer's decisions on the TITPs collected from L_1 and L_2 zones for different grades. Grad-CAM localizes the predicted class's activation maps and highlights the clinically significant regions for prediction. The areas having large gradient values contribute more to the final predicted score [33]. Saliency maps in Fig. 2 highlight the efficacy of the DiCoMLP-Mixer model in comprehending the micro-structural complex tissue convolutions manifested by the dense and compact regions of the TITPs and their relative significance towards final grading.

IV. Conclusion

This work examines the diagnostic efficacy of DiCoMLP-Mixer, a multi-scale context aggregation enriched MLP-mixer model for an exhaustive characterization of TITPs belonging to five study groups: healthy, OPMDs (OSF, OSFD, OLKP), and OC. Dilated convolutions in the patchify stem of DiCoMLP-Mixer abstract the computationally efficient multi-scale contextual feature abstraction by exponential receptive field expansion. At the same time, the MLP-mixer backbone in DiCoMLP-Mixer learns the saliency pertinent to particular sub-epithelial OCT grades. It leverages the attention mechanism of transformers in OMHIs by ensemble feature combination of the OMHI patches and the spatial encoding of these patches. The TITPs are obtained from the OMHIs acquired at two sub-epithelium zones of the OCT layers: papillary (L_1) and reticular (L_2). These sub-epithelium zoning accentuates the significance of depth-aware subtle structural variations with the onset of pre-cancerous activity. Grad-CAM heatmap localizes the saliency of the predicted class's activation and justifies DiCoMLP-Mixer's explainability for clinical deployment.

V. Acknowledgment

The authors acknowledge SMST and ATDC dept., of IIT Kharagpur and Guru Nanak Institute of Dental Science & Research Kolkata for providing research facility and IOP project (IIT/SRIC/MM/IOP/2017-18/208) of Higher education (Sci. and Tech.), Govt of West Bengal for research funding.

References

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 71, no. 3, pp. 209–249, 2021.

[2] [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/oral-health

[3] D. T. Jamison, A. Alwan, C. N. Mock, R. Nugent, D. Watkins, O. Adeyi, S. Anand, R. Atun, S. Bertozzi, Z. Bhutta *et al.*, "Universal health coverage and intersectoral action for health: key messages from disease control priorities," *The Lancet*, vol. 391, no. 10125, pp. 1108–1120, 2018.

[4] H. Amarasinghe, R. D. Jayasinghe, D. Dharmagunawardene, M. Attygalla, P. A. Scuffham, N. Johnson, and S. Kularatna, "Economic burden of managing oral cancer patients in sri lanka: a cross-sectional hospital-based costing study," *BMJ open*, vol. 9, no. 7, p. e027661, 2019.

[5] H. Mortazavi, M. Baharvand, and M. Mehdipour, "Oral potentially malignant disorders: an overview of more than 20 entities," *Journal of dental research, dental clinics, dental prospects*, vol. 8, no. 1, p. 6, 2014.

[6] M. M. R. Krishnan, V. Venkatraghavan, U. R. Acharya, M. Pal, R. R. Paul, L. C. Min, A. K. Ray, J. Chatterjee, and C. Chakraborty, "Automated oral cancer identification using histopathological images: a hybrid feature extraction paradigm," *Micron*, vol. 43, no. 2-3, pp. 352–364, 2012.

[7] M. Aubreville, C. Knipfer, N. Oetter, C. Jaremenko, E. Rodner, J. Denzler, C. Bohr, H. Neumann, F. Stelzle, and A. Maier, "Automatic classification of cancerous tissue in laserendomicroscopy images of the oral cavity using deep learning," *Scientific reports*, vol. 7, no. 1, pp. 1–10, 2017.

[8] J. Folmsbee, X. Liu, M. Brandwein-Weber, and S. Doyle, "Active deep learning: Improved training efficiency of convolutional neural networks for tissue classification in oral cavity cancer," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 770–773.

[9] P. R. Jeyaraj and E. R. Samuel Nadar, "Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm," *Journal of cancer research and clinical oncology*, vol. 145, no. 4, pp. 829–837, 2019.

[10] S. Xu, C. Liu, Y. Zong, S. Chen, Y. Lu, L. Yang, E. Y. Ng, Y. Wang, Y. Wang, Y. Liu *et al.*, "An early diagnosis of oral cancer based on three-dimensional convolutional neural networks," *IEEE Access*, vol. 7, pp. 158 603–158 611, 2019.

[11] B. Song, S. Sunny, R. D. Uthoff, S. Patrick, A. Suresh, T. Kolur, G. Keerthi, A. Anbarani, P. Wilder-Smith, M. A. Kuriakose *et al.*, "Automatic classification of dual-modalilty, smartphone-based oral dysplasia and malignancy images using deep learning," *Biomedical optics express*, vol. 9, no. 11, pp. 5318–5329, 2018.

[12] A. Rana, G. Yauney, L. C. Wong, O. Gupta, A. Muftu, and P. Shah, "Automated segmentation of gingival diseases from oral images," in *2017 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT)*. IEEE, 2017, pp. 144–147.

[13] R. Anantharaman, M. Velazquez, and Y. Lee, "Utilizing mask r-cnn for detection and segmentation of oral diseases," in *2018 IEEE international conference on bioinformatics and biomedicine (BIBM)*. IEEE, 2018, pp. 2197–2204.

[14] M. R. K. Mookiah, M. Pal, S. K. Bomminayuni, C. Chakraborty, R. R. Paul, J. Chatterjee, and A. K. Ray, "Automated classification of cells in sub-epithelial connective tissue of oral sub-mucous fibrosis - an svm based approach," *Computers in biology and medicine*, vol. 39 12, pp. 1096–104, 2009.

[15] M. M. R. Krishnan, P. Shah, M. Ghosh, M. Pal, C. Chakraborty, R. R. Paul, J. Chatterjee, and A. K. Ray, "Automated characterization of sub-epithelial connective tissue cells of normal oral mucosa: Bayesian approach," *2010 IEEE Students Technology Symposium (TechSym)*, pp. 44–48, 2010.

[16] M. R. K. Mookiah, P. Shah, C. Chakraborty, and A. K. Ray, "Statistical analysis of textural features for improved classification of oral histopathological images," *Journal of Medical Systems*, vol. 36, pp. 865–881, 2010.

[17] D. Nawn, S. Pratiher, S. Chattoraj, D. Chakraborty, M. Pal, R. R. Paul, S. Dutta, and J. Chatterjee, "Multifractal alterations in oral sub-epithelial connective tissue during progression of pre-cancer and cancer," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 1, pp. 152–162, 2020.

[18] M. M. R. Krishnan, V. Venkatraghavan, and C. Chakraborty, "Knowledge based segmentation, quantitative characterization and classification of basement membrane from oral histopathological images," *Journal of Medical Imaging and Health Informatics*, vol. 1, pp. 107–115, 2011.

[19] T. Y. Rahman, L. B. Mahanta, C. Chakraborty, A. K. Das, and J. D. Sarma, "Textural pattern classification for oral squamous cell carcinoma," *Journal of Microscopy*, vol. 269, 2018.

[20] M. M. R. Krishnan, U. R. Acharya, C. Chakraborty, and A. K. Ray, "Automated diagnosis of oral cancer using higher order spectra features and local binary pattern: A comparative study," *Technology in Cancer Research & Treatment*, vol. 10, pp. 443 – 455, 2011.

[21] M. M. R. Krishnan, P. Shah, A. Choudhary, C. Chakraborty, R. R. Paul, and A. K. Ray, "Textural characterization of histopathological images for oral sub-mucous fibrosis detection." *Tissue & cell*, vol. 43 5, pp. 318–30, 2011.

[22] M. R. K. Mookiah, C. Chakraborty, R. R. Paul, and A. K. Ray, "Hybrid segmentation, characterization and classification of basal cell nuclei from histopathological images of normal oral mucosa and oral submucous fibrosis," *Expert Syst. Appl.*, vol. 39, pp. 1062–1077, 2012.

[23] K. C. Figueroa, B. Song, S. P. Sunny, S. Li, K. Gurushanth, P. Mendonca, N. Mukhia, S. Patrick, S. Gurudath, S. Raghavan, T. Imchen, S. T. Leivon, T. C. Kolur, V. Shetty, V. Bushan, R. M. Ramesh, V. Pillai, P. B. B. Wilder-Smith, A. Sigamani, A. Suresh, M. A. Kuriakose, P. Birur, and R. Liang, "Interpretable deep learning approach for oral cancer classification using guided attention inference network," *Journal of Biomedical Optics*, vol. 27, 2022.

[24] C. Walker, E. Mojares, and A. del Río Hernández, "Role of extracellular matrix in development and cancer progression," *International journal of molecular sciences*, vol. 19, no. 10, p. 3028, 2018.

[25] D. A. Stewart, C. R. Cooper, and R. A. Sikes, "Changes in extracellular matrix (ecm) and ecm-associated proteins in the metastatic progression of prostate cancer," *Reproductive Biology and Endocrinology*, vol. 2, no. 1, pp. 1–13, 2004.

[26] P. Rajalalitha and S. Vali, "Molecular pathogenesis of oral submucous fibrosis–a collagen metabolic disorder," *Journal of oral pathology & medicine*, vol. 34, no. 6, pp. 321–328, 2005.

[27] K. S. Arora, A. Nayyar, P. Kaur, K. S. Arora, A. Goel, and S. Singh, "Evaluation of collagen in leukoplakia, oral submucous fibrosis and oral squamous cell carcinomas using polarizing microscopy and immunohistochemistry," *Asian Pacific journal of cancer prevention: APJCP*, vol. 19, no. 4, p. 1075, 2018.

[28] I. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, X. Zhai, T. Unterthiner, J. Yung, A. P. Steiner, D. Keysers, J. Uszkoreit *et al.*, "Mlp-mixer: An all-mlp architecture for vision," in *Thirty-Fifth Conference on Neural Information Processing Systems*, 2021.

[29] D. V. Dawson, D. R. Drake, J. R. Hill, K. A. Brogden, C. L. Fischer, and P. W. Wertz, "Organization, barrier function and antimicrobial lipids of the oral mucosa," *International journal of cosmetic science*, vol. 35, no. 3, pp. 220–223, 2013.

[30] F. Yu and V. Koltun, "Multi-Scale context Aggregation by Dilated Convolutions," 2016.

[31] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.

[32] M. M. R. Krishnan, C. Chakraborty, R. R. Paul, and A. K. Ray, "Quantitative analysis of sub-epithelial connective tissue cell population of oral submucous fibrosis using support vector machine," *Journal of Medical Imaging and Health Informatics*, vol. 1, pp. 4–12, 2011.

[33] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.