Assessment of the association of deep features with a polynomial algorithm for automated oral epithelial dysplasia grading

Adriano B. Silva*, Cléber I. de Oliveira†, Danilo C. Pereira§*, Thaína A. A. Tosta‡,
Alessandro S. Martins§, Adriano M. Loyola¶, Sérgio V. Cardoso¶,
Paulo R. de Faria‖, Leandro A. Neves† and Marcelo Z. do Nascimento*
*Faculty of Computer Science (FACOM), Federal University of Uberlândia (UFU)
†Department of Computer Science and Statistics (DCCE), São Paulo State University (UNESP)
‡Science and Technology Institute, Federal University of São Paulo (UNIFESP)
§Federal Institute of Triângulo Mineiro (IFTM)
¶Area of Oral Pathology, School of Dentistry, Federal University of Uberlândia (UFU)
‖Department of Histology and Morphology, Institute of Biomedical Science, Federal University of Uberlândia (UFU)
E-mails: *adriano.barbosa@ufu.br, cleber.oliveira@unesp.br, danilo.pereira@ufu.br*
tosta.thaina@unifesp.br, alessandro@iftm.edu.br, loyola@ufu.br, sv.cardoso@ufu.br
paulo.faria@ufu.br, leandro.neves@unesp.br, marcelo.nascimento@ufu.br

***Abstract*—Oral epithelial dysplasia is a potentially malignant lesion that presents challenges for diagnosis. The use of digital systems in histological analysis can aid specialists to obtain data that allows a robust and fast grading process, but there are few methods in the literature proposing a grading system for this lesion. This study presents a method for oral epithelial dysplasia grading in histopathological images combining deep features and a polynomial classifier. The ResNet50 and AlexNet models were trained with the images and information was extracted from the convolutional layers, exploring convolutional neural networks via transfer learning. Then, the ReliefF algorithm was used to rank and select the most relevant features, which were given as an input to the polynomial classifier. The methodology was employed in a dataset with 296 regions of mice tongue images. The results were compared with the gold standard and other algorithms present in the literature. The classification stage presented AUC values ranging from 0.9663 to 0.9800. When compared to other algorithms present in the literature, our method provided relevant results regarding accuracy and AUC values. The proposed approach presented relevant results and can be used as a tool to aid pathologists in grading oral dysplastic lesions.**

I. Introduction

Cancer is the second most common cause of death worldwide, with an estimation of 20 million new cases by about 2025. The World Health Organization estimates that appropriate treatment could avoid 7.3 million cancer deaths between 2020 and 2030 [1]. Oral cancer is one of the most common types of cancer and is the sixth leading cause of death in the world [2]. These aspects encourage the study of this type of lesion, which is gaining more attention with the proposed methods for diagnosis as a way to fasten its detection [3].

One way to reduce its frequency is the diagnosis and treatment of precocious lesions known as potentially malignant disorders, such as oral epithelial dysplasia (OED). OEDs present a potential for malignant transformation [3], with alterations to the shape, size and internal structures of cell nuclei. OEDs are graded as mild, moderate and severe [4]. One of the criteria to diagnose OED is the severity of epithelial cells' nuclear alterations while its grade takes into consideration the epithelium's thickness compromised by them [3], [5]. Early diagnosis of this lesion is important so that patients can receive the appropriate treatment based on its severity, reducing its risk of malignant transformation [6].

Diagnoses of such lesions are traditionally performed by microscopic analysis. With the large number of images that specialists have to analyze, this task leads to a repetitive routine that can be influenced by several factors, leading to misinterpretation and limitation in diagnosis accuracy. There are studies in the literature showing that intra and inter-observer divergences poses challenges in OED grading [3].

With the advance of technological resources and the digitization of medical images, computer-aided diagnosis (CAD) systems have emerged to assist specialists' decision-making diagnoses while analyzing such abnormalities. These systems allow the quantitative analysis of histological abnormalities in oral cavity images [7]. In such systems, steps are taken to improve the signal-to-noise ratio of the image, region detection, clinical finding segmentation, feature extraction and tissue classification [8]. An increase in the number of studies proposing the construction of CAD systems has been observed [9], most of them focused on identifying malignancies at advanced stages [10], [11], with different techniques being applied for feature extraction and histological image classification.

The use of deep learning networks for feature extraction allows the extraction of complex feature sets from the layers and hierarchical information can be obtained from different layers. As a result, deep learning models offer features that can

be used to efficiently represent images [12]. Even though several classification methods have been proposed, many studies have explored orthogonal polynomials for efficient solutions. This model can work with non-linearly separable classes and is an important solution for classification tasks. Hermite polynomials are relevant functions that use nonparametric density estimation and can be used for classification tasks [13].

In the literature, several studies were proposed for cancerous lesion detection, but few focused on automated grading of OEDs [14]. In the study presented by Baik et al. [11], lesions cancerous lesions were classified using the random forests (RF) algorithm, showing an accuracy rate of 80%. The proposed method was able to identify potentially malignant lesions, but no assessment regarding the definition of lesion grading was performed. Oral leukoplakia and squamous cell carcinoma images were classified by Banerjee et al. [15] using biomarkers extracted from infrared spectroscopy. These features were classified with the support vector machine (SVM) algorithm, achieving an accuracy value of 89.7%. This study focused on the identification of these lesions, but areas of malignant transformation were not investigated.

The work of Adel et al. [14] proposed a method for OED lesion identification combining the oriented FAST and rotated BRIEF (ORB) feature extractor with the SVM classifiers. The algorithm achieved an accuracy of 92.8% for the classification of OED tissues, but no assessment was made regarding the lesion grades. OED lesions were graded in the study of Silva et al. [16]. Handcrafted features of morphology, entropy and Jaccard index were extracted from segmented nuclei and given as input to a polynomial algorithm for classification. The authors achieved and A_{CC} average of 92,4%. The methodology was capable of identifying the OED grades, but only handcrafted features were employed in the study.

From the surveyed studies, it is noted that few works were focused on OED lesions. Only two of them diagnosed OED lesions and only one study performed the lesion grading process. Since OEDs show potential for a malignant evolution, it is important to grade oral dysplastic lesions because it allows the definition of the adequate treatment [3]. Based on the above-mentioned studies, it can be seen that this topic still faces challenges in improving the performance of systems in a classification of OED grades and that new research is needed so that further advances can be made with this approach. Moreover, this proposal is the first study to perform this task using deep learning features.

A. Contributions of this Work

This study proposes an approach for the automatic OED grading of histological images using deep features obtained through convolutional neural networks (CNN) models. Histological images were given as input to two CNN architectures to extract features from the images. A polynomial classifier was employed to classify the tissue images and the results obtained with features from the CNN models were compared. The main contributions of this study are summarized below:

- A new approach for feature extraction of OED tissue images based on CNN models regarding lesion grading;
- Analysis of the most relevant features and the impact of the number of features over the classification stage;
- Study of an association of deep features and the polynomial classifier capable of providing rates of distinction for different OED grades;
- Contribution to the state-of-the-art of automated methods for the identification and grading of precancerous lesions.

II. Materials and Methods

A. Image Dataset

The image dataset was built from 30 H&E-stained mice tongue tissue sections previously submitted to a carcinogen during two experiments carried out in 2009 and 2010. These experiments were approved by the Ethics Committee on the Use of Animals under protocol number 038/09 at the Federal University of Uberlândia, Brazil.

The histological slides were digitised using the Leica DM500 optical microscope with 400× magnification. A total of 66 images were obtained and stored in the TIFF format using the RGB colour model with a resolution of 2048×1536 pixels. Using the methodology described by [17], the images were classified between healthy mucosa, mild, moderate and severe OED by one specialist. From the images, 74 ROIs of size 450× 250 pixels were obtained for each class. Examples of these ROIs can be seen on Figure 1.

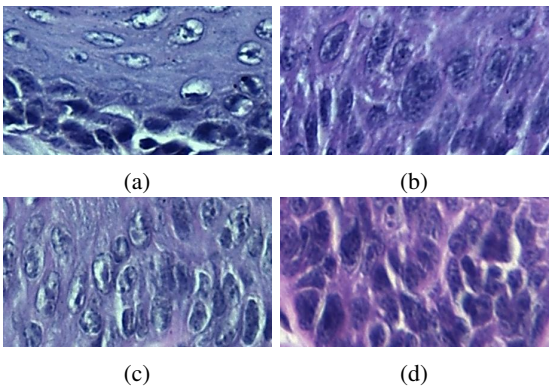


Fig. 1: Examples of oral histological tissues: (a) healthy tissue, (b) mild dysplasia, (c) moderate dysplasia and (d) severe dysplasia.

B. Feature Extraction

The image features were obtained using two CNN models. In these experiments, the convolutional and max-pooling layers before the flattening layer were computed. The outputs of flattening layer were stored in feature vectors.

The deep features were obtained via transfer learning of the AlexNet and ResNet50 architectures pre-trained on the ImageNet dataset. The AlexNet model consists of five convolutional layers interleaved with three poolings, two fully connected layers and a softmax function [18]. This architecture

is illustrated on Figure 2. The ResNet50 model is composed of an input layer followed by 48 convolutional layers arranged in repeated blocks, 16 skip connections and the classification layer, as can be seen on Figure 3. The initial layers extract features that quantify objects' edges, shapes and color. The final layers are used to identify global texture patterns. From the AlexNet model, 616,032 features were extracted from the five convolutional layers, excluding the fully connected and the softmax layers. From the ResNet50 architecture, a total of 1,095,488 features were obtained from the first two and the last three layers.

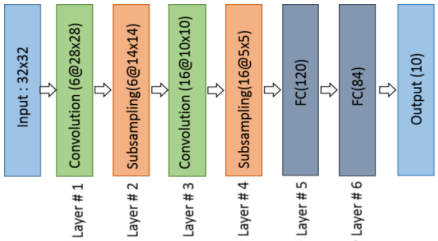


Fig. 2: Architecture of the AlexNet model used for extracting features of the tissue images [Source: Alom et al. [19]].

Every CNN layer was represented by n-dimensional matrices M_i, where i represents the layer. The columns of a M_i matrix were sequentially organized into a feature vector V_i. After creating the vectors, the most relevant features were obtained using the ReliefF algorithm [20]. Thereafter, each V_i was distributed in subsets S_m where m represents the number of features in each subset. Given the features obtained with each CNN model, an empirical investigation was performed based on the study of Ribeiro et al. [21]. Then, in this experiment, the 20 best-ranked features were selected for evaluation in the classification stage.

C. Classification

The feature vectors were classified using the polynomial algorithm proposed by Martins et al. [13], which employs Hermite orthogonal polynomial (HOP) functions. In the literature, methods based on polynomial functions are used on classification tasks due to their ability to work with non-linearly separable data and to identify the information about the lesions [22].

This classifier employs polynomial expansions over the feature vectors to define coefficients that allow the classification of samples. The discriminant function is obtained by:

$$g(x) = w_0 + \sum_{i=1}^{d} w_i x_i + \sum_{i=1}^{d}\sum_{j=1}^{d} w_{ij} x_i x_j, \tag{1}$$

where w_i is a vector of weights with a polynomial basis and x_i is a feature vector and d represents the number of features.

The polynomial discriminant functions can be generalized by:

$$g(x) = \sum_{i=0}^{L} a_i p_n(x_i) = a^T p_n(x), \tag{2}$$

where a is an L-dimensional weight vector, $p_n(x)$ is an L-dimensional vector whose entries are arbitrary functions of x and n is the order or degree of the polynomial.

With the value of $g(x)$, the data were assigned to the classes ω_1 or ω_2 using the decision rule defined by:

$$Decide \begin{cases} \omega_1, & \text{if } g(\mathbf{x}) < 0. \\ \omega_2, & \text{if } g(\mathbf{x}) > 0. \end{cases} \tag{3}$$

The Hermite orthogonal polynomial functions are defined by:

$$H_n(x) = (-1)^n e^{\frac{x^2}{2}} \frac{d^n}{dx^n} e^{-\frac{x^2}{2}}. \tag{4}$$

The feature vectors were passed as inputs defined by $\mathbf{x} = [x_1, x_2, \ldots, x_N]$ and expanded in terms of the polynomial basis vector $H_n(\mathbf{x})$. This process resulted in an L-dimensional vector.

Then, the polynomial expansions performed over N samples $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ were concatenated as:

$$M = [H_n(\mathbf{x}_1)\ H_n(\mathbf{x}_2)\ \ldots\ H_n(\mathbf{x}_N)]^T. \tag{5}$$

The problem proposed in Equation 3 was transformed into a matrix to determine the adequate coefficient vector for Equation 6.

$$Ma = b, \tag{6}$$

where the vector of coefficients is calculated by $a = (M^T M)^{-1} M^T b = M^{\dagger} b$, being the matrix $M^{\dagger}$ the inverse of M. The vector $b = [b_1, b_2, \ldots, b_n]^T$ is the ideal output with value -1 or 1.

The classifier output is computed by:

$$g(\mathbf{x}) = a^T H_n(\mathbf{x}), \tag{7}$$

where a is a coefficient vector, $H_n(\mathbf{x})$ is the Hermite basis function and n is the order of the polynomial function. These experiments were performed for each binary group separately, that is, healthy *vs* mild ($HvsM_i$), healthy *vs* moderate ($HvsM_o$), healthy *vs* severe ($HvsS$), mild *vs* moderate (M_ivsM_o), mild *vs* severe (M_ivsS) and moderate *vs* severe (M_ovsS). Based on the study of Karthik et al. [23] and due to the reduced number of images in the dataset, the algorithm was trained using a k-fold value of 5, resulting in 116 images for training and 29 for testing on each fold.

D. Evaluation

To assess the separability of the OED grades, the Mann-Whitney U test was computed for the feature vectors of each binary group. This nonparametric test was employed to quantify the significance of these features for the discrimination among the groups, with features having a p-value of 0.05 or less considered statistically significant.

The evaluation of the methodology was performed by comparing the obtained results with the gold standard classification performed by the specialist, in which the metrics of accuracy (A_{CC}) and area under the ROC curve (AUC) were obtained. The A_{CC} was used to measure the number of images correctly classified according to the gold standard. The AUC is a metric

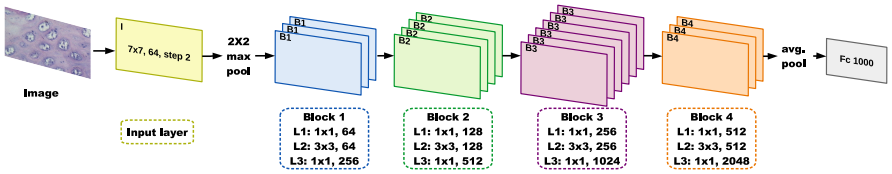


Fig. 3: Architecture of the ResNet50 model used in the feature extraction stage [Source: Silva et al. [16]].

extensively used to evaluate the performance of classifier algorithms [24], [25].

The HP classifier was, then, compared to other algorithms found in the literature, namely decision tree (DT), multilayer perceptron (MP), random forests (RF), ResNet50 and AlexNet. These algorithms were chosen based on the relevant results obtained in the classification task of histological images [26]–[28]. The MP algorithm is a neural network composed of the input and output layers and one or more hidden layers called perceptrons. Each layer contains neurons that use a nonlinear activation function [28]. The DT defines a tree structure with established decision rules and nodes that verify a certain condition. If the condition is satisfied, the flow goes through one branch, and if it is not satisfied, it goes through another branch [27]. The RF algorithm generates two or more trees with random decision rules. Every tree performs the classification task and outputs the obtained results, which are used to compute the average and define the final results [26]. At this stage, the Wilcoxon's signed-rank test was employed in order to evaluate the statistical relevance between the HOP and the other machine learning algorithms.

III. Results and Discussion

In this section, the results obtained with the proposed methodology are presented and discussed. The classification was performed using the features obtained from the AlexNet and ResNet50 architectures and the junction features from these models.

The separability significance analysis with the Mann-Whitney U test is shown in Figure 4. Figure 4a shows the features extracted from the ResNet50 model. It is observed that the $HvsM_i$ had the lowest value of separable features, around 75%. All the other groups showed more than 80% of separable features, with the M_ivsS and the M_ovsS groups having around 95%. The assessment of the features extracted with the AlexNet model are shown in Figure 4b. It is observed that the M_ovsS group presented a significant value of 100% number of separable features. However, the other groups presented lower values than the ResNet50, with most of them presenting values around 70%, and the M_ivsM_o group showing a value of 60%. From the features obtained by combining the information obtained from the two models, shown in Figure 4c, it was observed that the number of separable features was lower than the other feature sets. The $HvsM_i$ group presented a value of around 45% and the M_ivsM_o group showed a value of 55%.

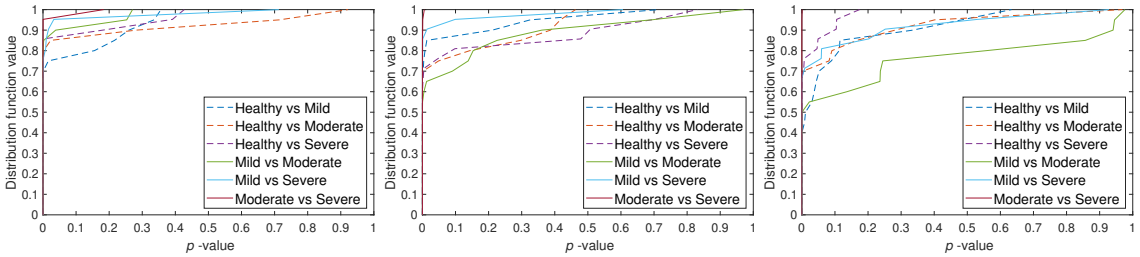
The comparison among these feature sets is shown in Table I and the average AUC and A_{CC} obtained for the binary combinations are presented for each polynomial order. It was observed that the performance of the algorithm using the AlexNet features achieved the lowest results, with AUC values ranging from 0.9266 to 0.9570. The highest results were achieved by the ResNet50 features, with values ranging from 0.9663 to 0.9800. The values obtained using the junction features were slightly lower than the values obtained with the ResNet50. Moreover, the best values were obtained when using a polynomial expansion of second order. Considering that the results obtained by ResNet50 were more relevant than the other feature sets, thereafter all the analyzes and experiments were explored with this model.

TABLE I: Classification results obtained with the HOP algorithm for the different feature sets. Each line shows the average AUC obtained for each polynomial order.

Order	ResNet50	AlexNet	Junction features
1	0.9766 ± 0.0046	0.9568 ± 0.0065	0.9739 ± 0.0042
2	**0.9800 ± 0.0049**	**0.9570 ± 0.0049**	**0.9800 ± 0.0046**
3	0.9780 ± 0.0025	0.9484 ± 0.0068	0.9766 ± 0.0046
4	0.9663 ± 0.0074	0.9266 ± 0.0095	0.9665 ± 0.0044

Table II shows the results obtained with the HOP classifier, where each line represents the values achieved for the binary combinations of the OED grades. This experiment was performed using the Hermite polynomial expansion of second order because it achieved results more relevant than the other orders. For the $HvsM_i$, $HvsM_o$ and $HvsS$ groups, the HOP classifier achieved $AUC = 1$. On the M_ivsM_o comparison, the classifier showed an AUC value of 0.9365, the lowest value achieved by the classifier. For the groups M_ivsS and M_ovsS, the AUC values obtained were of 0.9838 and 0.9597, respectively. It is observed that the lowest values were presented in comparisons containing the moderate class.

The classifiers' results from the literature are also shown in Table II. For all algorithms, the findings involving healthy tissues presented higher results than the other OED grades. The MP algorithm showed AUC values higher than 0.90 for all comparisons, with the exception of the M_ivsM_o, which achieved a value of 0.8449. The RF classifier presented values slightly higher than those of the MP, with AUC of 0.9976 and 0.8812 for the $HvsS$ and M_ivsM_o groups, respectively. The DT classifier achieved the lowest results on the M_ivsM_o and M_ovsS groups, showing values of 0.7272 and 0.7816, respectively. The ResNet50 model achieved results slightly lower than the HOP, with an AUC values of 0.9932 for the $HvsM_i$ and 0.9842 for the $HvsM_o$. This CNN model achieved the highest value for the group M_ivsM_o, with an



(a) ResNet50 (b) AlexNet (c) Merged features

Fig. 4: Empirical cumulative distribution function of the p-value of features: (a) ResNet50, (b) AlexNet and (c) merged features.

$AUC = 0.9437$. The AlexNet presented AUC values ranging from 0.9146 to 0.9831.

Among all classifiers, the highest values were achieved by the HOP classifier, obtaining an average AUC of 0.9800, followed by the ResNet50 with average $AUC = 0.9706$. The AlexNet model ranked third, with an average AUC of 0.9615. The p-values obtained through the Wilcoxon's test indicate that the results achieved by the HOP and the Resnet50 are so close that there is no statistical significance between them. However, the proposed method showed relevant improvement when compared to the other classifier methods.

Table III shows a comparison between the obtained results and the relevant computer vision methods for tissue classification on oral lesions-derived histopathological images. It is important to highlight that the work by Silva et al. [16] used the same dataset as this study and extracted handcrafted features for the classification stage. Moreover, the methods described by Das et al. [29] and Baik et al. [11] were developed to distinguish healthy tissues from advanced stage lesions. It is also noted that the study presented by [14] investigated OED images, but no assessment regarding their grades was proposed. Considering these results, our proposal provides values compatible with those available in the literature and achieves significant results for the grading process. These results are important for reducing the subjectivity in grading OED at the time of diagnosis.

IV. Conclusion

This work presented an approach for OED grading based on deep features in combination with a HOP classifier capable of providing rates of distinction for different OED grades. The proposed methodology presented a relevant result of an average AUC of 0.98. The minimum value was of 0.9365 compared to the mild and moderate grades. All comparisons with healthy tissues achieved $AUC = 1$.

The obtained results indicate that the features extracted from the ResNet50 model are better suited for the classification of OED images than those extracted by AlexNet or the junction features from both CNN models. The values obtained with the AUC metric showed that the association of the deep features and the HOP classifier can achieve significant results for OED classification. In the experiments, the HOP classifier presented higher values than the other algorithms and the highest AUC values achieved with this classifier were using a Hermite polynomial expansion of second order.

In the literature, there are few studies proposing computational methods for OED grading and this methodology contributes to the state-of-the-art. The presented results can be used to help pathologists during the histological analysis of OED lesions. Future studies will explore the impact of data augmentation on the methodology and the combination of handcrafted and deep features. Other classification algorithms, such as Naive Bayes and CNN models, will be employed in combination with the deep features to investigate the impact of such classifiers. Furthermore, other datasets of histological images will be employed to assess the efficacy of the methodology for general histopathology diagnosis.

Acknowledgment

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. The authors gratefully acknowledge the financial support of National Council for Scientific and Technological Development CNPq (Grants #313643/2021-0 and #311404/2021-9), the State of Minas Gerais Research Foundation - FAPEMIG (Grant #APQ-00578-18 and Grant #APQ-01129-21).

References

[1] World Health Organization, "Who report on cancer: setting priorities, investing wisely and providing care for all," 2020.

[2] 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," in *Students' Technology Symposium (TechSym), 2010 IEEE*. IEEE, 2010, pp. 44–48.

[3] D. Kademani, *Improving Outcomes in Oral Cancer: A Clinical and Translational Update*. Springer Nature, 2019.

[4] T. Fonseca-Silva, M. G. Diniz, S. F. Sousa, R. S. Gomez, and C. C. Gomes, "Association between histopathological features of dysplasia in oral leukoplakia and loss of heterozygosity," *Histopathology*, vol. 68, no. 3, pp. 456–460, 2016.

[5] V. Kumar, J. C. Aster, and A. Abbas, *Robbins and Cotran Patologia-Bases Patológicas das Doenças*. Elsevier Brasil, 2010.

[6] V. Kumar, A. K. Abbas, and J. C. Aster, *Robbins patologia básica*, 9th ed. Elsevier Brasil, 2013.

[7] A. Belsare and M. Mushrif, "Histopathological image analysis using image processing techniques: An overview," *Signal & Image Processing*, vol. 3, no. 4, p. 23, 2012.

[8] R. C. Gonzalez and R. Woods, "Digital image processing," 2018.

TABLE II: Comparison of the results obtained by the classifiers. Each line shows the AUC obtained for the binary combinations.

Sets	Classifiers					
	HOP	**MP**	**DT**	**RF**	**ResNet50**	**AlexNet**
$HvsM_i$	**1.0000 ± 0.0000**	0.9984 ± 0.0039	0.9409 ± 0.0427	0.9977 ± 0.0040	0.9932 ± 0.0027	0.9831 ± 0.0079
$HvsM_o$	**1.0000 ± 0.0000**	0.9995 ± 0.0026	0.9412 ± 0.0438	0.9993 ± 0.0019	0.9842 ± 0.0048	0.9784 ± 0.0083
$HvsS$	**1.0000 ± 0.0000**	0.9909 ± 0.0124	0.9794 ± 0.0280	0.9976 ± 0.0054	0.9821 ± 0.0071	0.9811 ± 0.0068
M_ivsM_o	0.9365 ± 0.0184	0.8449 ± 0.0647	0.7272 ± 0.0928	0.8812 ± 0.0459	**0.9437 ± 0.0144**	0.9146 ± 0.0194
M_ivsS	**0.9838 ± 0.0060**	0.9610 ± 0.0266	0.8751 ± 0.0477	0.9644 ± 0.0280	0.9775 ± 0.0096	0.9653 ± 0.0105
M_ovsS	**0.9597 ± 0.0051**	0.9127 ± 0.0652	0.7816 ± 0.0835	0.9208 ± 0.0682	0.9429 ± 0.0103	0.9462 ± 0.0081
Average	**0.9800 ± 0.0049**	0.9512 ± 0.0292	0.8742 ± 0.0564	0.9602 ± 0.0256	0.9706 ± 0.0082	0.9615 ± 0.0102
p-value	-	0.03	0.03	0.03	0.15	0.03

TABLE III: Comparison of methods used by the related works.

Methods	**Dataset images**	**Pre-processing**	**Segmentation**	**Feature Extracted**	**Classification**	**Accuracy**
Das et al. [29]	Total of 90 images. 30 for each grade of SCC	Bit-plane slicing and contrast normalisation	Active contour snake	Keratin regions	Percentage of keratin present in the image	95.08%
Baik et al. [11]	Total of 133 images. 29 healthy, 40 with carcinoma in situ and 64 with SCC	—	Manual segmentation by specialists	Morphology, DNA amount, chromatin distribution	Random Forests	80.6%
Adel et al. [14]	Total of 138 images. 21 healthy and 117 dysplastic	Conversion to HSV colour model	—	ORB, SIFT	SVM	92.8%
Silva et al. [16]	Total of 296 images. 74 for each OED grade	—	Mask R-CNN combined with morphological operations	Morphology, entropy, Moran's index	HOP	92.4%
Proposed	Total of 296 images. 74 for each OED grade	—	—	Deep features obtained with the AlexNet and ResNet50 models	HOP	**98.0%**

[9] M. J. Huttunen, R. Hristu, A. Dumitru, I. Floroiu, M. Costache, and S. G. Stanciu, "Multiphoton microscopy of the dermoepidermal junction and automated identification of dysplastic tissues with deep learning," *Biomedical Optics Express*, vol. 11, no. 1, pp. 186–199, 2020.

[10] M. M. R. Krishnan, 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 Systems with Applications*, vol. 39, no. 1, pp. 1062–1077, 2012.

[11] J. Baik, Q. Ye, L. Zhang, C. Poh, M. Rosin, C. MacAulay, and M. Guillaud, "Automated classification of oral premalignant lesions using image cytometry and random forests-based algorithms," *Cellular Oncology*, vol. 37, no. 3, pp. 193–202, 2014.

[12] D. M. Vo, N.-Q. Nguyen, and S.-W. Lee, "Classification of breast cancer histology images using incremental boosting convolution networks," *Information Sciences*, vol. 482, pp. 123–138, 2019.

[13] A. S. Martins, L. A. Neves, P. R. de Faria, T. A. Tosta, L. C. Longo, A. B. Silva, G. F. Roberto, and M. Z. do Nascimento, "A hermite polynomial algorithm for detection of lesions in lymphoma images," *Pattern Analysis and Applications*, vol. 24, no. 2, pp. 523–535, 2021.

[14] D. Adel, J. Mounir, M. El-Shafey, Y. A. Eldin, N. El Masry, A. AbdelRaouf, and I. S. Abd Elhamid, "Oral epithelial dysplasia computer aided diagnostic approach," in *2018 13th International Conference on Computer Engineering and Systems (ICCES)*. IEEE, 2018, pp. 313–318.

[15] S. Banerjee, M. Pal, J. Chakrabarty, C. Petibois, R. R. Paul, A. Giri, and J. Chatterjee, "Fourier-transform-infrared-spectroscopy based spectral-biomarker selection towards optimum diagnostic differentiation of oral leukoplakia and cancer," *Analytical and bioanalytical chemistry*, vol. 407, no. 26, pp. 7935–7943, 2015.

[16] A. B. Silva, A. S. Martins, T. A. A. Tosta, L. A. Neves, J. P. S. Servato, M. S. de Araújo, P. R. de Faria, and M. Z. do Nascimento, "Computational analysis of histological images from hematoxylin and eosin-stained oral epithelial dysplasia tissue sections," *Expert Systems with Applications*, p. 116456, 2022.

[17] H. Lumerman, P. Freedman, and S. Kerpel, "Oral epithelial dysplasia and the development of invasive squamous cell carcinoma," *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, vol. 79, no. 3, pp. 321–329, 1995.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[19] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, B. C. Van Esesn, A. A. S. Awwal, and V. K. Asari, "The history began from alexnet: A comprehensive survey on deep learning approaches," *arXiv preprint arXiv:1803.01164*, 2018.

[20] I. Kononenko, M. Robnik-Sikonja, and U. Pompe, "Relieff for estimation and discretization of attributes in classification, regression, and ilp problems," *Artificial intelligence: methodology, systems, applications*, pp. 31–40, 1996.

[21] M. G. Ribeiro, L. A. Neves, M. Z. do Nascimento, G. F. Roberto, A. S. Martins, and T. A. A. Tosta, "Classification of colorectal cancer based on the association of multidimensional and multiresolution features," *Expert Systems With Applications*, vol. 120, pp. 262–278, 2019.

[22] L. C. Padierna, M. Carpio, A. Rojas-Domínguez, H. Puga, and H. Fraire, "A novel formulation of orthogonal polynomial kernel functions for svm classifiers: the gegenbauer family," *Pattern Recognition*, vol. 84, pp. 211–225, 2018.

[23] R. Karthik and R. Menaka, "A multi-scale approach for detection of ischemic stroke from brain mr images using discrete curvelet transformation," *Measurement*, vol. 100, pp. 223–232, 2017.

[24] C. L. Chen, A. Mahjoubfar, L.-C. Tai, I. K. Blaby, A. Huang, K. R. Niazi, and B. Jalali, "Deep learning in label-free cell classification," *Scientific reports*, vol. 6, p. 21471, 2016.

[25] B. Q. Huynh, H. Li, and M. L. Giger, "Digital mammographic tumor classification using transfer learning from deep convolutional neural networks," *Journal of Medical Imaging*, vol. 3, no. 3, p. 034501, 2016.

[26] S. Graham, M. Shaban, T. Qaiser, N. A. Koohbanani, S. A. Khurram, and N. Rajpoot, "Classification of lung cancer histology images using patch-level summary statistics," in *Medical Imaging 2018: Digital Pathology*, vol. 10581. International Society for Optics and Photonics, 2018, p. 1058119.

[27] S. Alinsaif and J. Lang, "Texture features in the shearlet domain for histopathological image classification," *BMC Medical Informatics and Decision Making*, vol. 20, no. 14, pp. 1–19, 2020.

[28] M. Desai and M. Shah, "An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (mlp) and convolutional neural network (cnn)," *Clinical eHealth*, vol. 4, pp. 1–11, 2021.

[29] D. K. Das, C. Chakraborty, S. Sawaimoon, A. K. Maiti, and S. Chatterjee, "Automated identification of keratinization and keratin pearl area from in situ oral histological images," *Tissue and Cell*, vol. 47, no. 4, pp. 349–358, 2015.