Adaptive Segmentation of Oral Cavity Lesions via Color-based Custom Index and Differential Evolution

Manilo Monaco
Dept. of Information Engineering, Electrical Eng. and Applied Mathematics
University of Salerno
Salerno, Italy
mmonaco@unisa.it

Sabrina Senatore*
Dept. of Information Engineering, Electrical Eng. and Applied Mathematics
University of Salerno
Salerno, Italy
ssenatore@unisa.it
*Corresponding author

Mario G. C. A. Cimino
Dept. of Information Engineering
University of Pisa
Pisa, Italy
mario.cimino@unipi.it

Olga Di Fede
Dept. of Surgical, Oncological and Stomatological Studies
University of Palermo
Palermo, Italy
olga.difede@unipa.it

Abstract*— This work proposes an adaptive method for segmenting oral cavity lesions, designed to support AI-assisted diagnosis and improve early detection. The method is based on three basic components: (i) a segmentation index, inspired by the well-known Soil-Adjusted Vegetation Index (SAVI). In this work, the index is reinterpreted as a lesion index, exploiting subtle color variations between healthy and pathological tissues to identify clinically relevant areas of a lesion; (ii) a superpixel algorithm, which groups neighboring pixels into color-homogeneous regions to highlight its clinical morphology, and (iii) an evolutionary optimization algorithm, which tunes the parameters of both the index and the segmentation process to the specific clinical context. By leveraging intrinsic color features as diagnostic indicators, the approach enables robust segmentation even in the absence of annotated data. Preliminary experiments conducted on a representative set of oral lesion images demonstrate the effectiveness of the method. Segmentation performance was quantitatively evaluated using metrics such as Kappa coefficient, precision, showing promising results for clinical applications.

***Keywords*— image segmentation; oral cavity; simple linear iterative clustering; evolutionary algorithm**

I. Introduction and Background

Early diagnosis of oral cavity disorders, including potentially malignant and malignant lesions, is essential for improving patient outcomes and informing therapeutic strategies. Advancements in artificial intelligence (AI) and computer vision have progressively supported oral medicine professionals, achieving automated image analysis and segmentation workflows. However, the development of robust AI systems depends on the high-quality annotated datasets, which represent the main challenge in medical image analysis. The interpretation and manual annotation of oral cavity disorders' clinical images can vary significantly among practitioners, affecting the quality of the final ground truth annotation that in turn informs image segmentation [1]. Therefore, despite physicians follow a wide range of universally accepted clinical features [2], the identification of objective, reproducible criteria based on lesion colors and morphological features is of paramount importance.

A recent trend in the design of computer vision and deep learning-based pipelines for early detection of oral cancer involves the use of image pre-processing and segmentation steps [3]. The aim can be twofold: both minimize the impact of noise to improve models' capability to identify essential features within images and extract interpretable features to enable complex models to be more explainable. Deep learning models [4] have demonstrated promising results in segmenting oral potentially malignant lesions by effectively fusing features from adjacent image regions. This fusion enhances boundary delineation and captures subtle lesion variations, improving segmentation accuracy and robustness despite variability in clinical images. An alternative approach [5] combines quantum cascade laser-based discrete frequency infrared chemical imaging with a CNN for multi-class tissue segmentation, showing potential for improved diagnosis of oral cavity disorders.

Although annotation protocols are based on widely accepted clinical criteria [2], the lack of objective and reproducible segmentation standards often compromises ground truth quality and then, model performance. To address these challenges, segmentation methods based on color, texture, and morphology analysis have attracted increasing interest. For instance, in [6], fluorescence spectroscopy is combined with machine learning classifiers to detect and classify oral mucosal lesions, leveraging spectral color information from tissue autofluorescence. Another study [7] shows that color intensity-based textural features from PAP-stained cytologic images can distinguish oral squamous cell carcinoma from normal cells, demonstrating the potential of color-texture analysis for early screening.

This work proposes an adaptive methodology for segmenting oral cavity lesions that explicitly leverages the color characteristics of oral tissues, which serve as valuable indicators for identifying clinically relevant regions. Specific color patterns can correspond to pathologically altered areas that require closer inspection. An ad-hoc lesion index has been defined to quantitatively capture the chromatic properties of the lesion tissue, enabling better discrimination of color variations associated with pathological changes. This color-focused index is integrated with a superpixel-based segmentation framework and an evolutionary algorithm for optimizing segmentation parameters. The proposed method not only enhances segmentation accuracy but also extracts interpretable features that align closely with expert annotations, thereby advancing the development of explainable AI applications in oral medicine.

II. Materials and Methods

In this work, an adaptive method for segmentation of oral cavity lesions is presented. The method is based on three basic components: (i) a mathematical index, used to identify areas of clinical interest of a lesion, (ii) a superpixel algorithm, used to segment a lesion into color-homogeneous regions while highlighting its clinical morphology, and (iii) an evolutionary algorithm, used to comprehensively adapt the method parameters to the specific clinical needs of representation. The following subsections describe the dataset used in this study, the individual components involved in the proposed methodology, how the problem was defined, and the related evaluation metrics used for the parametric adaptation.

A. The Oral case study

The dataset used in this study consists of images of oral ulcerated lesions. It initially comprises 221 plain photographic images of clinical cases, collected following histological confirmation when necessary. These images were captured using a digital camera or a smartphone by the Oral Medicine Staff, including consultants, dental hygienists, and students, between 2015 and 2020 at the Oral Medicine Unit of the University Hospital P. Giaccone in Palermo, Italy.

Since the proposed method leverages the intensity of lesion colors and no color correction techniques were applied, only a few images were considered for the experimental phase to avoid biases related to the heterogeneity of the technical specifications of the different sensors used. After analyzing the metadata related to the model of the digital cameras used for images acquisition, it was decided to include only the samples acquired with the digital camera Nikon Coolpix S6 and the digital camera Nikon D7200, as they amount to more than half of the initial dataset and they are distributed over all lesion categories considered. Table I summarizes the total number of acquired images grouped by lesion category and the number of samples considered for the experimental phase, in turn partitioned by acquisition tool.

TABLE I. Images acquired by lesion category

Lesion Category	***Total***	***Coolpix S6***	***D7200***
	# of samples	***# of samples***	***# of samples***
Aphthous	61	48	3
Neoplastic	79	9	55
Traumatic	81	53	4
All	221	110	62

The images were annotated using the COCO Annotator annotation tool. Lesions were manually annotated by a trained practitioner, with each lesion defined by a bounding polygon and an associated label. Bounding boxes were generated automatically by the tool based on the extreme coordinate points of the annotated polygons. To ensure consistency and accuracy, all annotations underwent a review process carried out by a senior dentist.

B. The Soil-Adjusted Vegetation Index

A preliminary analysis was carried out on the entire image dataset to extract useful information on the color features of the lesions. Fig. 1 shows the distributions of the mean and standard deviation of color (Red, Green, and Blue) intensity levels for the foreground and background pixels, corresponding to the pixels inside and outside the polygon bounding a lesion, respectively. The histograms reveal that red intensity levels tend to be higher in areas inside lesions. Moreover, for all three-color channels, the histograms show less variability in intensity levels within lesion areas.

The histograms shown in Fig. 1 also highlight an abrupt transition in the mean spectral reflectance levels between the blue or green band and the visible red band, probably due to the increased blood flow in the superficial mucosa of the areas affected by the lesions. This behavior exhibits similarities with the "red edge", a fundamental phenomenon exploited in vegetation observation by remote sensing techniques, due to the combination of selective absorption of chlorophyll in the red band and the strong reflection and scattering of near-infrared light by leaf cells [8]. Therefore, to identify regions of clinical interest within an oral cavity lesion, the original idea was to borrow the analytical expression of a well-known spectral index directly related to the "red edge" effect, i.e., the Soil-Adjusted Vegetation Index (SAVI), as a lesion index. The SAVI was selected for its easy calculation and interpretation. It was successfully used in environmental monitoring and precision agriculture, typically computed on satellite multi-spectral or hyper-spectral images. To the best of our knowledge, this is the first time that a vegetation index has inspired the design of a lesion index in medical images.

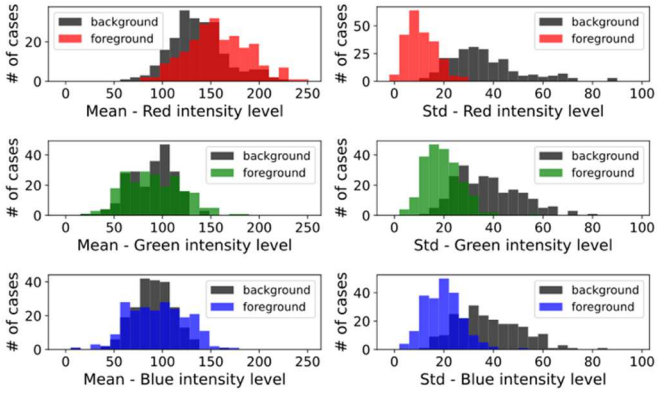

Fig. 1. Statistics of RGB colors intensity levels

The SAVI was developed to address limitations of the Normalized Difference Vegetation Index (NDVI) in areas with sparse vegetation cover. Specifically, SAVI mitigates the influence of soil brightness, a significant factor that can distort vegetation assessment in regions where vegetation density is low. Let us assume NIR and Red as the reflectance values in the Near-InfraRed and Red bands respectively, given an expert-provided factor L, the SAVI is computed according to the following formula:

$$SAVI = \frac{(1+L)(NIR-Red)}{(NIR+Red+L)} \quad (1)$$

In Equation (1) the parameter L is used to adjust for soil background effects and is commonly set to 0.5, balancing sensitivity to both vegetation and soil reflectance in mixed-pixel scenarios. In the original context, lower SAVI values indicate reduced vegetation cover, while higher values correspond to increased vegetation density. Compared to NDVI, SAVI provides a more reliable assessment of vegetation in areas with sparse or absent vegetation, making it particularly useful for arid and semi-arid regions where soil exposure significantly affects spectral measurements [9]. By reducing soil-induced variability, SAVI enhances the precision of vegetation monitoring and is widely applied in studies involving vegetation health, land degradation, and ecological assessment.

In our index formulation of the lesion index, the NIR value was replaced by the value of the radiance reflected in the Red band, while the Red value is replaced by the value of the radiance reflected in the Green or Blue band, based on adaptation of the evolutionary algorithm. Moreover, the parameter L can be interpreted as an attenuation factor of the brightness induced by the camera flash on saliva. In fact, the brightness due to the flash tends to result in a saturation of the intensity levels of the three bands and thus reduce the differences. In contrast, a higher L factor tends to amplify the differences. So, the factor L was also treated as a free parameter, and its choice was delegated to the Differential Evolution algorithm.

C. The Simple Linear Iterative Clustering

In the context of interpreting clinical images of oral lesions, a crucial aspect is the visual perception of the texture of the lesion and the immediately surrounding tissues. Therefore, to be able to correctly segment the chromatic and morphological changes inside and outside the lesion is a fundamental step in identifying objective criteria for clinical evaluation. In this work, the segmentation task was carried out by means of a superpixel algorithm, i.e., the Simple Linear Iterative Clustering (SLIC).

Superpixel algorithms partition an image into coherent and perceptually meaningful regions that serve as atomic units of representation, moving beyond the fixed structure of the pixel grid. These regions effectively capture image redundancy, simplify the extraction of image features, and significantly reduce the computational complexity of downstream processing. As a result, superpixels have become fundamental components in numerous computer vision methodologies. However, the SLIC has been shown to outperform modern superpixel techniques, in terms of boundary adherence, segmentation speed, and performance when used as a preprocessing step in a segmentation framework [10].

The SLIC algorithm is an efficient and widely used method for superpixel generation. It operates by clustering pixels in a five-dimensional space defined by their spatial coordinates and color values. The algorithm takes two main parameters: k and m, which play critical roles in controlling the segmentation process. The parameter k represents the approximate number of desired superpixels in the segmented image. Based on this, SLIC divides the image into k roughly uniform grid regions and initializes cluster centers within these regions. These centers are then iteratively refined to better align with the local structure of the image. The parameter m controls the relative importance of color similarity and spatial proximity in the clustering process. A higher m value emphasizes spatial proximity, leading to more compact superpixels, while a lower m value prioritizes color similarity, potentially producing irregularly shaped regions.

At each iteration, SLIC reassigns pixels to the nearest cluster center by minimizing a distance measure defined in the five-dimensional space of spatial and color information. The distance between a pixel $p = (x, y, l, a, b)$ and a cluster center $C = (x_c, y_c, l_c, a_c, b_c)$ is computed as:

$$D = \sqrt{\left(\frac{d_c}{N_c}\right)^2 + \left(\frac{d_s}{N_s}\right)^2} \tag{2}$$

where $d_c = \sqrt{(l - l_c)^2 + (a - a_c)^2 + (b - b_c)^2}$ represents the color distance in the CIELAB color space, and $d_s = \sqrt{(x - x_c)^2 + (y - y_c)^2}$ represents the spatial distance. N_c and N_s are normalization terms, with $N_s = S$, the side length of the square region initially assigned to each superpixel, and $N_c = m$, the weight parameter that adjusts the relative influence of color and spatial distance. The iterative process alternates between updating cluster centers as the mean of all assigned pixels and reassigning pixels to the nearest center until convergence. The algorithm guarantees compact and well-defined superpixels while maintaining computational efficiency.

In our method, k and m parameters were allowed to be selected by the evolutionary algorithm, to adapt the morphology of the segments to the annotations carried out by the clinical experts.

D. The Differential Evolution Algorithm

Recently, novel techniques based on genetic algorithms have been applied for automated composition of unsupervised image clustering algorithms, achieving new state-of-the-art performance results for oral lesion datasets [11]. In this paper, the final goal is quite different, i.e., to segment the bounding box of an oral lesion and compare the segmentation result with an annotation validated by a clinical specialist, in order to extract interpretable texture features. However, despite our features extraction technique was defined a priori by relying on the composition of two pillars, i.e., the SAVI and the SLIC, the combination of parameters changes in the single components critically impacts the overall capability to capture and isolate features as perceived by the clinical expert. Therefore, in this work the Differential Evolution (DE) algorithm, a type of evolutionary algorithm, was used to tune the parameters of the SAVI and the SLIC in a unique and continuous search space.

Evolutionary algorithms (EAs) are population-based metaheuristic optimization algorithms, based on computational mechanisms of biological evolution, such as reproduction, mutation, recombination, and selection of solutions. EAs can tackle non-linear and complex optimization problems, requiring just the objective function values. Nevertheless, the performance of an EA depends, in turn, on hyper-parameter settings: for instance, probabilities of mutation and crossover,

population size, and number of generations. Here, the DE algorithm has been considered, i.e., an evolutionary algorithm for optimization in continuous spaces. The performance of the DE depends on the mutation control hyper-parameters, especially when the problem is complex. To balance the convergence (fitness evaluations) and the reliability (optimum's globality), ranges of hyper-parameters values have been studied in other work [12]. A popular strategy of DE, as it is a good solution for many systems, is called "DE/best/1/bin", where "DE" stands for Differential Evolution, "best" means that the perturbing individual is generated from the best population member, "1" is the number of pairs of individuals chosen for mutation, and "bin" denotes the binomial crossover.

E. The problem statement

Formally, the objective function to be minimized is summarized by the pseudocode presented in Algorithm 1.

It is worth noting that the binary lesion mask is generated by thresholding the SAVI-based lesion index (LI) values, aggregated over each superpixel, in turn detected via the SLIC algorithm. In other words, for each superpixel, a single representative value is used to determine its classification as lesion or non-lesion. As a decision value, the statistical median of the LI values distribution was chosen, to reduce the impact of local outliers and noise.

Table II reports the parameter space and the related ranges or sets of values explored by the DE algorithm to optimize the objective function. These parameters directly correspond to the components of the solution x in the objective formulation.

TABLE II. PARAMETERS SPACE OF THE SAVI-SLIC-DE ALGORITHM

Parameters	Interval
b_1	$\{RED\}$
b_2	$\{BLUE, GREEN\}$
l	$[0, 1]$
n_s	$[8, 80]$
c	$[0.1, 10]$
t_l	$[-1, 1]$
t_u	$[-1, 1]$

The first two parameters, b_1 and b_2, are categorical variables representing spectral bands used in the computation of the SAVI-based lesion index, together with the adjustment parameter l. The first band (b_1) is fixed to RED to preserve the semantic consistency of the lesion index formulation. The ranges for parameters n_s and c, related to the SLIC segmentation, were selected to encourage the aggregation of wide and homogeneous color regions and to promote the generation of irregularly shaped superpixels, respectively. Both of which are favorable in delineating lesion contours. However, the final bounds were refined through preliminary experiments aimed at identifying the regions of the parameter space where optimal solutions tend to cluster. The segmentation thresholds (t_l, t_u) were constrained within the domain of the SAVI index definition. An additional constraint, i.e. $t_l < t_u$, was imposed on the optimizer to ensure semantic coherence in the thresholding process.

Algorithm 1

SAVI-SLIC Objective Function

Input:

- $x = (b_1, b_2, l, n_s, c, t_l, t_u)$
  - b_1, b_2: Bands for lesion index calculation
  - l: Adjustment parameter for the SAVI index
  - n_s: Number of segments of SLIC
  - c: Compactness for SLIC
  - t_l, t_u: Thresholds for binary segmentation
- $annotations_{df}$: DataFrame containing image filenames
- $images_{dir}, masks_{dir}$: Directories of images and reference masks

Output:

- Fitness value $f(x)$, to be minimized

Procedure:

1. Initialize the evaluation metric list

2. For each image in the dataset:

2.1. Load the original image and its corresponding mask

2.2. Resize both to $(600,800)$

2.3. Compute the Lesion Index (LI) using SAVI:

2.3.1. Select bands b_1, b_2

2.3.2. Compute LI using the formula:

$$LI = \frac{(1 + l) \cdot (B_{b_1} - B_{b_2})}{B_{b_1} + B_{b_2} + l}$$

2.4. Apply SLIC segmentation:

2.4.1. Perform SLIC with n_s segments and compactness c

2.4.2. Aggregate LI values per superpixel

2.5. Binary thresholding:

2.5.1. Generate a binary image:

$$img_{binary}(i,j) = \begin{cases} 1, if\, t_l < LI(i,j) \leq t_u \\ 0, otherwise \end{cases}$$

2.6. Compute the evaluation metric:

2.6.1. Convert reference mask to binary using Otsu's method

2.6.2. Compute weighted precision score using Cohen's Kappa:

$$f(x) = -Precision \times Cohen'sKappa$$

2.6.3. Append computed metric to the metric list

3. Return the mean of the computed metrics as the final fitness value

To assess the quality of lesion segmentation in oral cavity images, a custom evaluation metric was introduced. It was defined as the precision weighted by the Cohen's Kappa coefficient. Thus, for each individual x, the overall fitness value assessed by DE algorithm is computed according to the following formula:

$$fitness = \frac{1}{s} \cdot \sum_{i=1}^{s} \frac{TP}{TP+FP} \cdot \frac{p_o - p_e}{1 - p_e} \quad (3)$$

where s is the number of images within the lesion category; TP and FP represent the number of true positives and false positives, respectively; p_o is the observed agreement between the predicted segmentation and the ground truth, and p_e is the expected agreement by chance.

This composite score was designed to capture both local accuracy, i.e. how precise the model is in identifying lesion pixels, and global agreement, i.e. how well the overall segmentation agrees with expert annotations, beyond chance. Precision quantifies the proportion of predicted lesion pixels that are actually correct, penalizing over-segmentation and false alarms. This is particularly important in clinical applications where false positives can lead to unnecessary concern or follow-up examinations. Cohen's kappa provides a more holistic view of agreement, accounting for both lesion and non-lesion classes, and correcting for potential class imbalance.

The combination of precision and Cohen's kappa ensures a balanced and clinically meaningful evaluation. Although traditional metrics such as Dice or IoU are widely used, they can be insensitive to false positives or may not reflect agreement beyond chance. Precision emphasizes the need for conservative and trustworthy lesion detection. Cohen's Kappa ensures the model captures the expert-defined lesion patterns in a statistically robust manner. This composite metric thus aligns both with the clinical expectations of specificity and the statistical reliability of expert agreement.

III. Experimental Results and Discussion

The experiments were performed on a virtual machine based on x86_64 architecture, managed through the KVM hypervisor. The system has two virtual CPUs (vendor: Intel Corp.), each with 8 physical cores, for a total of 16 cores, without active hyper-threading (1 thread per core). The clock frequency of each core is 2.0 GHz. The first-level cache (L1) is 512 KiB for each core (for both data and instructions), the L2 cache is 64 MiB in total (presumably 4 MiB per core), and the L3 cache amounts to 32 MiB overall. The system is equipped with 64 GiB of RAM, distributed across four virtual modules of 16 GiB each, with ECC (Error-Correcting Code) support enabled. During the experiments, approximately 60 GiB of free memory was available, indicating a controlled environment free from interference from other concurrent processes.

For each category of lesion and digital camera used, the experimental phase was conducted on a subset of cases from the original dataset, also to reduce processing time. The number of cases considered compared to the total number of cases available is shown in Table IV. Since the metrics used to evaluate segmentation quality are sensitive to unbalanced datasets, cases with the best balance between foreground and background pixels were considered for the experiment. The average percentages of balance between the two-pixel classes are also shown in Table IV. The performance values reported in Table IV were obtained by considering the confidence intervals for the mean value of the selected metrics on three independent runs of Differential Evolution. For each lesion category and digital camera, the combination of parameters and thresholds that determined the best fitness following optimization with DE is shown in Table III.

Fig. 2 illustrates some representative examples of oral cavity lesion segmentations. In particular, the following are shown: a) the original images (bounding boxes), b) the annotations of the lesions made by specialized medical personnel, c) the segmentations corresponding to the combinations of parameters and thresholds relating to the individual who recorded the median performance in the initial generation, i.e., before optimization with DE, d) the segmentations corresponding to the combinations of parameters and thresholds relating to the individual with the best fitness in the last generation. The experimental findings confirm that the proposed adaptive segmentation approach achieves robust performance across diverse lesion types and imaging conditions. Notably, the integration of the Differential Evolution algorithm allows the system to dynamically adjust to variations in lesion appearance and capture modality, yielding segmentations that exhibit high consistency with expert clinical annotations. These results suggest the method's strong potential for supporting diagnostic workflows in heterogeneous clinical settings. Fig. 3 illustrates, for each lesion category, the average fitness of individuals across generations, recorded during the optimization process that allowed the identification of the best-performing combination of parameters and thresholds, shown in Table III. The optimization dynamics driven by the DE algorithm exhibit a progressive convergence toward parameter configurations associated with higher fitness values. This behaviour reflects a more efficient and targeted exploration of the search space, in which the algorithm quickly abandons suboptimal initial regions and concentrates its search on areas more likely to yield clinically coherent segmentations. The observed trend indicates that DE effectively guides the optimization toward meaningful parameter sets, reducing redundant evaluations and accelerating convergence toward solutions that better match expert-derived ground truth.

A. Limitations

Despite the promising results, some limitations should be acknowledged when assessing the generalizability and clinical applicability of the proposed methodology. The experimental dataset is relatively limited in size and composition. The initial dataset included 221 clinical images collected between 2015 and 2020, but the experimental phase was conducted on a reduced subset of samples to ensure acquisition homogeneity, minimize biases related to specific camera characteristics, and reduce computational time due to the hardware constraints. Although this choice enabled consistent analysis, it limited the overall variability observed in the data and may limit the ability of the methodology to generalize to acquisition conditions not represented in this study (e.g., different devices, illumination settings, shooting angles, or different demographic characteristics).

TABLE III. BEST COMBINED PARAMETERS AND THRESHOLDS AFTER DE-BASED OPTIMIZATION

Lesion category	Digital camera	Best SAVI parameters			Best SLIC parameters		Segmentation thresholds	
		Band 1	*Band 2*	*L*	*n. segments*	*compactness*	*Lower thresh*	*Upper thresh*
Neoplastic	Nikon Coolpix S6	RED	BLUE	0.98510848	12	7.78800742	0.25755591	0.33218003
	Nikon D7200	RED	GREEN	0.14916162	19	3.80107807	0.1679243	0.31738033
Aphthous	Nikon Coolpix S6	RED	GREEN	0.94040795	9	2.33698795	-0.06789947	0.3199806
	Nikon D7200	RED	GREEN	0.84703889	21	0.78349002	0.11304727	0.37168375
Traumatic	Nikon Coolpix S6	RED	GREEN	0.69374776	20	7.70090192	-0.90396221	0.19789148
	Nikon D7200	RED	BLUE	0.06982729	18	0.13818026	0.28278597	0.52036195

TABLE IV. PERFORMANCE EVALUATION AFTER DE-BASED OPTIMIZATION

Lesion category	Digital camera	# of cases	FG/BG Balance	Confidence Intervals for Mean (using t-Student distribution) over 3 DE runs			
				Precision (P)	*Cohen's Kappa (K)*	*Optimized Custom Metric (P * K)*	*DE Time (Hours)*
Neoplastic	Nikon Coolpix S6	9/9	0.60341	0.75823 ± 0.00106	0.26404 ± 0.01380	0.21917 ± 0.01034	48.99 ± 31.03
	Nikon D7200	9/55	0.50745	0.74320 ± 0.01145	0.35669 ± 0.01533	0.28176 ± 0.01075	80.90 ± 23.82
Aphthous	Nikon Coolpix S6	9/48	0.63803	0.92352 ± 0.00439e-03	0.43202 ± 0.02852e-03	0.41935 ± 0.02151e-03	36.02 ± 3.50
	Nikon D7200	3/3	0.68510	0.91122 ± 0.00885	0.37206 ± 0.02371	0.33871 ± 0.02051	21.87 ± 7.21
Traumatic	Nikon Coolpix S6	9/53	0.53836	0.79023 ± 0.02755e-05	0.33310 ± 0.06315e-05	0.28105 ± 0.02804 e-05	76.05 ± 5.53
	Nikon D7200	4/4	0.56215	0.74979 ± 0.17148	0.30995 ± 0.02631	0.25986 ± 0.04492	52.98 ± 38.23

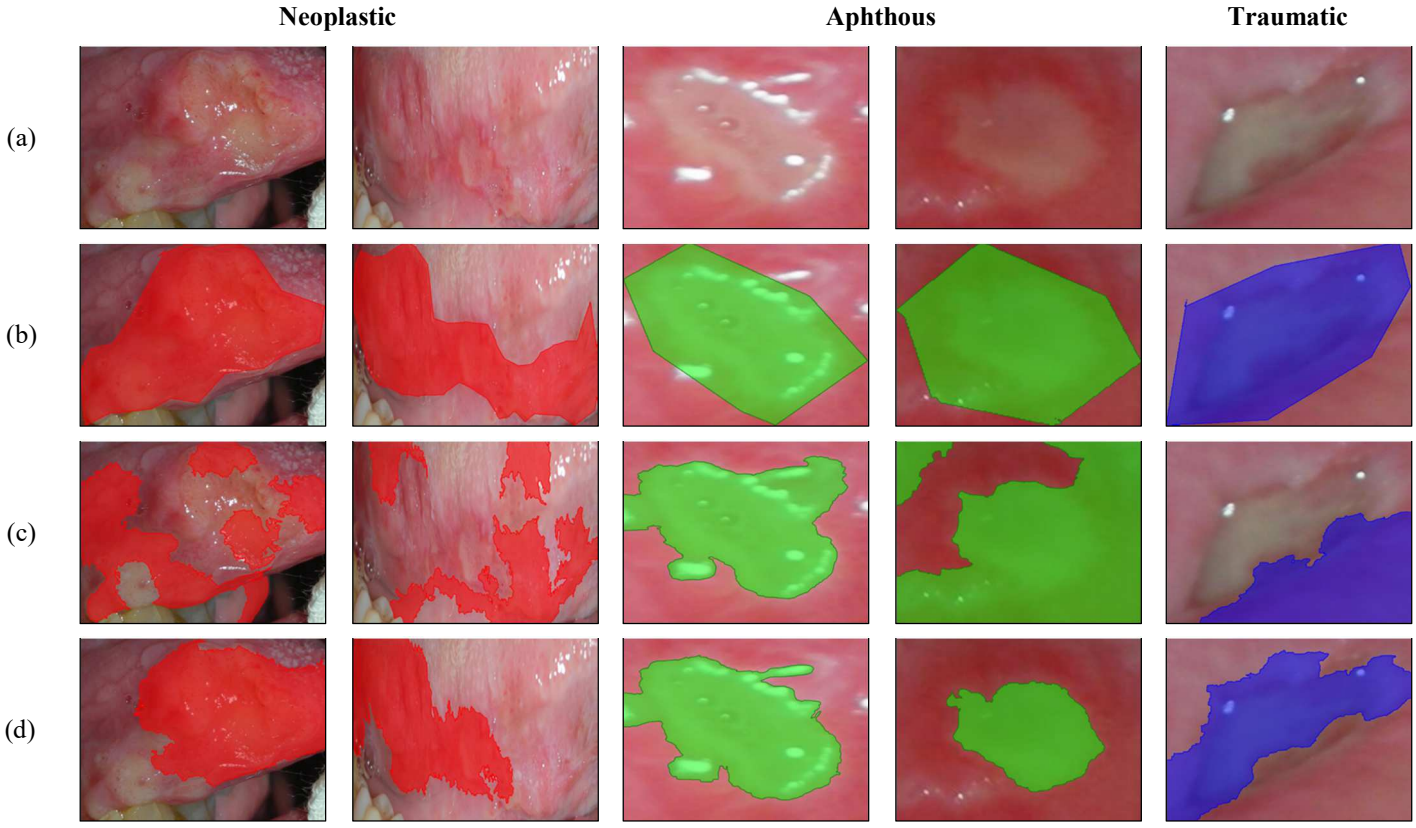

Fig. 2. Examples of annotations: (a) original bounding box, (b) Expert's annotation, (c) Before DE-based optimization, (d) After DE-based optimization

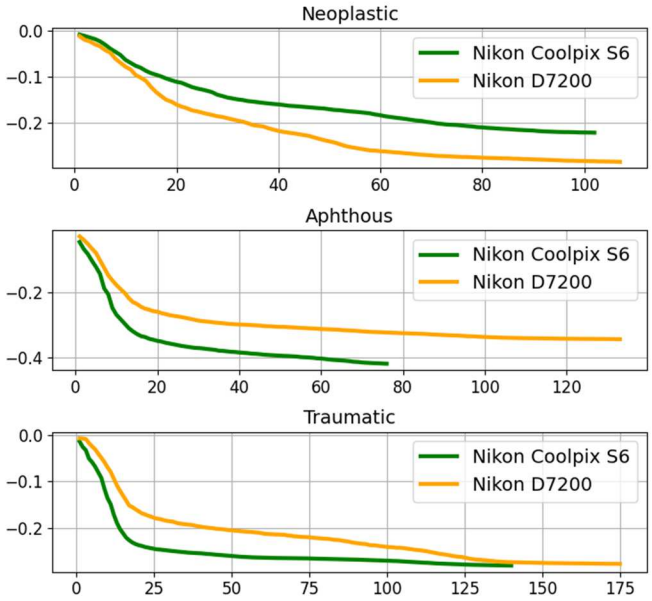

Fig. 3. Custom metric optimization: average best fitness vs number of generations

As highlighted in the literature, small and medium-sized clinical datasets can lead to inflated performance estimates and selection biases; therefore, the results presented in this study should be considered preliminary and subject to validation in larger multicenter cohorts [13].

Furthermore, the clinical scenarios considered are restricted. The dataset originates from a single clinical unit and includes only selected lesion categories (aphthous, neoplastic, traumatic). The effectiveness of the methodology on other pathological conditions, atypical lesions, or different populations remains untested. Demonstrating clinical robustness will require to incorporate multicenter datasets and real-world acquisitions (multiple operators, mobile devices, variable ambient illumination). Moreover, the methodology is affected by color variability issues. The absence of standardized colorimetric corrections and the use of flash introduce systematic variations in color rendering (saturation, white balance variations, gamma differences). Comparative studies on the color measurements of photographic devices have shown how camera quality and settings affect color accuracy. Without standardized calibration procedures (e.g., color reference targets, white-balance correction, color profiles) systematic errors in color measurements may occur, potentially undermining the reliability of color indices derived from RGB channels [14].

The proposed lesion index (LI) while computationally simple and interpretable, presents intrinsic drawbacks. Its design, inspired by the soil-adjusted vegetation index (SAVI), originally designed for multispectral images with near-infrared (NIR) information, was adapted here to RGB images. Although this adaptation enhances interpretability, it carries limitations: (i) low-complexity indices may saturate or lose sensitivity under strong illumination changes or in tissues with atypical chromatic properties; and (ii) simplified parameterization (e.g., a single adjusted factor L) may not capture all clinical variability sources. Previous studies on vegetation indices have documented saturation effects and background sensitivity, suggesting caution in transferring these indices to non-original domains [15].

Finally, the methodology has not yet been prospectively validated in real clinical workflows, and its contribution to decision-making and diagnostic performance remains to be assessed.

IV. Conclusions

This paper introduces a novel, adaptive methodology for segmenting oral cavity lesions. By explicitly leveraging the color characteristics of oral tissues through a lesion index, integrated with a superpixel-based segmentation framework and an evolutionary algorithm for optimization, our approach moves beyond the limitations of subjective manual annotations. The adaptation of the vegetation index to the clinical domain provides a novel and effective way to identify color variation associated with the pathology changes in the clinical images, offering a further measure to improve the segmentation.

Our preliminary experimental results demonstrate the effectiveness of this adaptive method across different lesion categories and camera types. The Differential Evolution algorithm ensures the methodology aligns closely with expert clinical perception. The proposed composite evaluation metric, combining precision with Cohen's Kappa, provides a balanced and clinically relevant assessment of segmentation quality, emphasizing both local accuracy and global agreement.

By extracting interpretable features that align with clinical expert knowledge, our method enhances segmentation accuracy and fosters trust in diagnostic workflows of oral cavity disorders. Future work will focus on expanding the dataset and further validating the method's performance in other clinical scenarios. The proposed segmentation strategy could be integrated into enhanced AI workflows for object detection and classification [16], enriching input data with additional channels or supporting semi-supervised and incremental learning scenarios. Such extensions may improve both data efficiency and clinical interpretability in AI-based diagnostic systems.

Acknowledgments

Work partially supported by: (i) the European Commission under the NextGenerationEU program, Extended Partnership PNRR PE1 - "FAIR - Future Artificial Intelligence Research" - Spoke 1 "Human-centered AI", and PNRR - M4 C2, Investment 1.5 "Creating and strengthening of "innovation ecosystems", building "territorial R&D leaders", project "THE - Tuscany Health Ecosystem", Spoke 6 "Precision Medicine and Personalized Healthcare"; (ii) the Italian Ministry of Education and Research (MIUR) in the framework of the FoReLab project (Departments of Excellence), in the framework of the "Reasoning" project, PRIN 2020 LS

Programme, Project number 2493 04-11-2021, and in the framework of the project "OCAX -Oral CAncer eXplained by DL-enhanced case-based classification" PRIN 2022 code P2022KMWX3; (iii) the Italian Ministry of Enterprises and Made in Italy , in the framework of the "Agreements for Innovation" Projet "4DDS - 4D Drone Swarms" Ref. no. F/310097/01-04/X56. (iv) the Tuscany Region in the framework of the WAU project, PR FESR 2021-2027, Project No. 27716.29122023.042000115. (v) the Digital Republic Fund supported by Google.org in the framework of the wAIne project, CrescerAI national call, Project No. 2023-CRE-00288.

REFERENCES

[1] Araújo, Anna Luíza Damaceno, et al. "Clinicians' perception of oral potentially malignant disorders: a pitfall for image annotation in supervised learning." Oral surgery, oral medicine, oral pathology and oral radiology 136.3 (2023): 315-321.

[2] Warnakulasuriya, Saman, et al. "Oral potentially malignant disorders: a consensus report from an international seminar on nomenclature and classification, convened by the WHO Collaborating Centre for Oral Cancer." Oral diseases 27.8 (2021): 1862-1880.

[3] Devindi, G. A. I., et al. "Multimodal deep convolutional neural network pipeline for AI-assisted early detection of oral cancer." IEEE Access (2024).

[4] Yang Ouyang, Yongxiu Du, Yuefu Zhan, Yinwei Zhan, Shanglei Gao, Zijun Chen, "AAFNet: An oral potentially malignant disorder image segmentation network based on adjacent feature fusion", Journal of Radiation Research and Applied Sciences, Volume 18, Issue 1, 2025, 101265, ISSN 1687-8507, https://doi.org/10.1016/j.jrras.2024.101265.

[5] Confer, Matthew P., Kianoush Falahkheirkhah, Subin Surendran, Sumsum P. Sunny, Kevin Yeh, Yen-Ting Liu, Ishaan Sharma, Andres C. Orr, Isabella Lebovic, William J. Magner, and et al. 2024. "Rapid and Label-Free Histopathology of Oral Lesions Using Deep Learning Applied to Optical and Infrared Spectroscopic Imaging Data" Journal of Personalized Medicine 14, no. 3: 304. https://doi.org/10.3390/jpm14030304.

[6] Kumar P, Rathod S, Pradhan A. Detection of oral mucosal lesions by the fluorescence spectroscopy and classification of cancerous stages by support vector machine. Lasers Med Sci. 2024 Jan 19;39(1):42. doi: 10.1007/s10103-024-03995-3. PMID: 38240832.

[7] Sharma PN, Chaudhary M, Patel SA, Zade PR. Screening of Oral Squamous Cell Carcinoma Through Color Intensity-Based Textural Features. Cureus. 2024 Mar 22;16(3):e56682. doi: 10.7759/cureus.56682. PMID: 38646364; PMCID: PMC11032690.

[8] Y. Sun, Q. Qin, H. Ren, T. Zhang and S. Chen, "Red-Edge Band Vegetation Indices for Leaf Area Index Estimation From Sentinel-2/MSI Imagery," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 2, pp. 826-840, Feb. 2020, doi: 10.1109/TGRS.2019.2940826.

[9] D. Cavaliere and S. Senatore, "Multi-grained wildfire damage estimation from satellite vegetative scenario by fuzzy decision tree," *2022 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Padua, Italy, 2022, pp. 1-8, doi:10.1109/FUZZ-IEEE55066.2022.9882730.

[10] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 11, pp. 2274-2282, Nov. 2012, doi: 10.1109/TPAMI.2012.120.

[11] Mia Gerber and Nelishia Pillay. 2023. Automated algorithm composition of unsupervised image clustering algorithms. In Proceedings of the Companion Conference on Genetic and Evolutionary Computation (GECCO '23 Companion). Association for Computing Machinery, New York, NY, USA, 467–470. https://doi.org/10.1145/3583133.3590555.

[12] Mario G.C.A. Cimino, Domenico Minici, Manilo Monaco, Stefano Petrocchi, Gigliola Vaglini, "A hyper-heuristic methodology for coordinating swarms of robots in target search", Computers and Electrical Engineering, Volume 95, 2021, 107420, ISSN 0045-7906, https://doi.org/10.1016/j.compeleceng.2021.107420.

[13] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Embracing imperfect datasets: A review of deep learning solutions for medical image analysis with limited data," *Medical Image Analysis*, vol. 63, p. 101693, Apr. 2020, doi: 10.1016/j.media.2020.101693.

[14] R. Saincher, A. Sholapurkar, R. D. D. Dias, S. Jayaraman, S. Kulkarni, and P. G. Tandon, "Comparison of color accuracy and picture quality of digital DSLR, point-and-shoot, and mobile cameras used for intraoral photography," *Journal of Dental Anesthesia and Pain Medicine*, vol. 22, no. 5, pp. 303–312, Oct. 2022, doi: 10.17245/jdapm.2022.22.5.303.

[15] Z. Zhen, Y. Xu, X. Li, J. Liu, G. Chen, and W. Wang, "Using the Negative Soil Adjustment Factor in Vegetation Indices Based on the Red and Near-Infrared Spectral Bands," *Sensors*, vol. 21, no. 12, p. 3995, Jun. 2021, doi: 10.3390/s21123995.

[16] Mario G.C.A. Cimino, Giuseppina Campisi, Federico A. Galatolo, Paolo Neri, Pietro Tozzo, Marco Parola, Gaetano La Mantia, Olga Di Fede, Explainable screening of oral cancer via deep learning and case-based reasoning, Smart Health, Volume 35, 2025, 100538, ISSN 2352-6483, https://doi.org/10.1016/j.smhl.2024.100538.