

Image Processing Methods for Oral Macules and Spots Segmentation

Carolina R. Kelsch¹, Jean Schmith¹, Rita F. T. Gomes²,
Vinicius C. Carrard^{2,3}, Rodrigo Marques de Figueiredo¹

¹Polytechnic School – University of Vale do Rio dos Sinos (UNISINOS)
São Leopoldo 93022-750 – RS – Brazil

²Department of Oral Pathology, Faculdade de Odontologia
Federal University of Rio Grande do Sul (UFRGS)
Porto Alegre 90035-003 – RS – Brazil

³TelessaúdeRS, Federal University of Rio Grande do Sul (UFRGS)
Porto Alegre 91501-970, Brazil

carol-kelsch@hotmail.com, j.schmith@gmail.com, marquesf@unisinós.br

Abstract. *Oral cancers are the 16th most common type of cancer in the world and present a high mortality rate. This is mainly because they are frequently discovered in an advanced stage due to the lack of specialized professionals. Some clinical characteristics such as borders and symmetry can aid in cancer diagnosis, and therefore the segmentation of the lesions is important. In light of this, this work aimed to present and evaluate different analytic methods to perform automatic segmentation of oral macules and spots from 21 clinical images. From the tested methods, the one with the best result reached an accuracy of 84.9%, a precision of 70.1%, a recall of 75.3%, and an f1-score of 60.8%, which are similar outcomes of published works that used artificial intelligence.*

1. Introduction

Computer vision and artificial intelligence are increasingly being used to automate and accelerate analysis procedures in the health field. Many studies have been developed in the area of lesion classification by images, but there is a paucity of studies in the area of stomatology. In the last decade, more works have been published in the area of automated classification and detection of oral cancer, but there are still very few that perform lesion segmentation and, with the advancement of technology, there is room for new approaches.

Mouth cancer is the most common type of head and neck cancer in the world. According to GLOBOCAN analysis reports, the projected age-standardized worldwide incidence rate for lip and oral cavity cancer in 2018 is 4.0 per 100,000, and the age-standardized mortality rate was 2.0 per 100,000 individuals [Sarode et al. 2020]. Overall, mouth cancer accounts for 3 percent of all malignancies in men and 2 percent in women in the U.S., with an increased risk observed with increasing age. According to the latest GLOBOCAN 2018 report, mouth cancer is the nineteenth most common cancer in America [Sarode et al. 2020], and have been associated with alcohol consumption, smoking, HPV infection for cancers of the oropharyngeal region, and ultraviolet radiation from exposure to sunlight for lip cancer [Sung et al. 2021]. Although early diagnosis reduces

morbidity and mortality, diagnoses are predominantly made at an advanced stage, for several reasons, including the scarcity of professionals sufficiently trained to recognize early-stage lesions due to the absence of symptoms and the heterogeneity in the appearance of oral lesions, which leads to delays in referring patients to specialists [Tanriver et al. 2021].

The identification of Potentially Malignant Disorders (PMD) is critical to improve early detection of oral cancer and, has inspired the development of screening tools for this disease [Tanriver et al. 2021]. Automatic segmentation of lesions with malignant potential or defining the boundaries between malignant lesions and healthy tissue can contribute to the definition of peripheral areas where the lesion is more diffuse, and consequently difficult to assess precisely, even for experienced specialists [Song et al. 2022]. Lesion segmentation is already widely applied in the field of dermatology and aids in the classification of cancerous lesions [dos Santos et al. 2022]. And just as in the classification of skin lesions, asymmetry and borders are important markers of aggressive cancers such as oral melanomas. In addition, lesion segmentation can assist in tracking and evaluating lesion progression such as PMDs.

To the best of our knowledge, only three studies performed segmentation of oral lesions in clinical images. One of them used texture features with a BPANN network (Backpropagation Artificial Neural Network) and obtained an accuracy of 97.9%, which, despite having been tested with few images, indicates that the texture has a high discriminating power [Thomas et al. 2013]. Another work used a Mask R-CNN (Mask Region-based Convolutional Neural Network), which has a higher complexity than BPANN, and obtained an accuracy of 74.4% [Anantharaman et al. 2018]. Finally, there is also a paper that used an U-Net for segmentation and obtained a 92.6% dice score using EfficientNet-b7 as the basis. The Mask R-CNN was also tested and the best result found an average accuracy of 37.98% [Tanriver et al. 2021]. One of the difficulties stated in the articles on cancer detection by images is the lack of open databases with image data of oral lesions [Tanriver et al. 2021] and the poor standardization of the existing in terms of quality from the images [Welikala et al. 2020], [Gomes et al. 2023a], [Gomes et al. 2023b]. Most works make use of histopathological, radiological, and hyperspectral images [Mahmood et al. 2021], but these images require a higher acquisition technology and do not address the problem of facilitating the follow-up of lesions on a larger scale of the population.

Therefore, this study developed and analyzed three proposals for segmenting oral lesions using pixel-intensity features in clinical images. The work focused only on the segmentation of oral macules and spots. The segmentation proposals in this work did not use machine learning, maintaining a low complexity with a focus on a low processing requirement.

2. Methodology

In this work, different image segmentation techniques and methods were applied and compared for the delineation of macules and spots. The purpose was to create a mask that extracted only the lesion, keeping the highest possible fidelity. To this end, three lesion segmentation methods were defined. The methods were tested with 21 images provided by TelessaúdeRS, "a research center linked to the post-graduation program in epidemiology at the Medical School from the Federal University of Rio Grande do Sul

(UFRGS)” [TelessáudeRS 2022]. Among the actions of TelessáudeRS, EstomatoNet is a telediagnostic service for diseases that manifest themselves as lesions in the mouth. The images provided by TelessáudeRS follow the ethical guidelines of the Declaration of Helsinki [World Medical Association 2013]. They represent 15 different lesions, classified as macules or spots. The images are not standardized, presenting different angles, illumination, focus, sharpness, and ratio between lesion and background. The image with the highest resolution have 987x1123 pixels, while the lowest have 88x76 pixels.

2.1. Image Pre-processing and Ground Truth Definition

The images were resized to present more similar dimensions, with the largest dimension adopting 320 pixels and the smallest a multiple of 32 pixels that kept at most the original proportion. After that, the ground truths of the lesions were generated, a mask that delimits the segmentation region of the lesion considered to be true. This was accomplished with the aid of superpixel segmentation, which groups several pixels with similar characteristics into large pixels, reducing the complexity of the image. The MSLIC, or Manifold SLIC [Liu et al. 2016], algorithm was used, which has the characteristic of creating larger superpixels in sparse regions with less information, and smaller superpixels in the regions of the image with more information. Thus, the pixels containing the lesion were manually selected and the ground truth of each of the images was formed. The selection was supervised by the two specialist authors. Figures 1a and 1b presents an example of the superpixel mask generated for a lesion and the ground truth of this lesion.

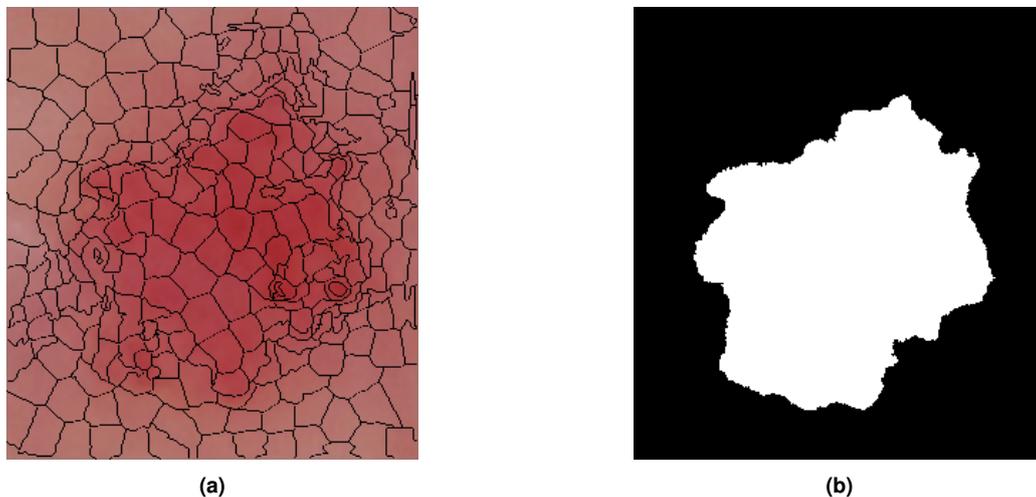


Figure 1. Ground truth generation process, in (a) the superpixels image and in (b) the ground truth of the lesion, generated by manual selection of superpixels.

2.2. Lesion Segmentation Methods

To perform the segmentation of the lesions, three different methods were tested, using the grayscale image. In addition, a normalization of the gray intensities from 0 to 255 was performed in order to improve the contrast.

2.2.1. Segmentation by Otsu's Thresholding

The first method consisted in segmenting the lesions by Otsu's thresholding [Otsu 1979], applied on the grayscale inverted images, obtaining a binary mask where the lesion is represented by the white area. An example of the process can be seen in Figure 2.

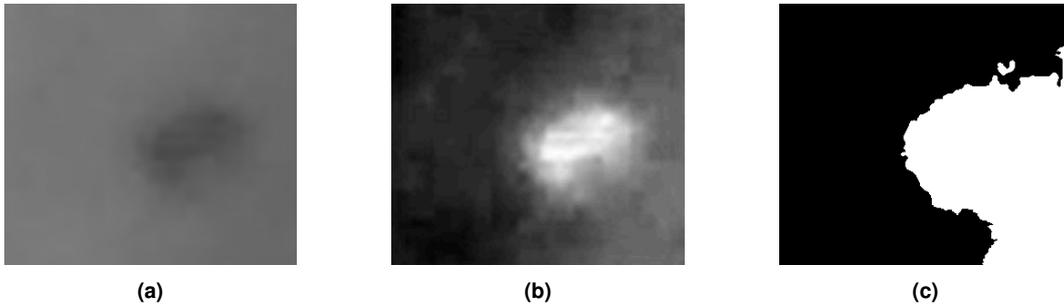


Figure 2. Otsu's thresholding segmentation process, in (a) the grayscale image from the lesion, in (b) the image normalized and inverted and in (c) the binarized image by Otsu's thresholding.

2.2.2. Segmentation by Otsu's Thresholding on Backprojection Images

The second method applied the backprojection technique, widely used in the reconstruction of images from CT scans. This method involves the creation of an image from several captures made at different angles. The projection from each angle was here accomplished by rotating the original image and summing the pixels in the same column. After that, the projected image is reoriented by rotating it at the same angle as the original, but in the opposite direction. At the end of the process, these images are summed. In Figure 3b, the result of summing the 45° spaced projections of Figure 3a is shown, which is known as the backprojection image. Figure 3c presents the result of summing the projections of the image from 0° to 179° with steps of 1° .

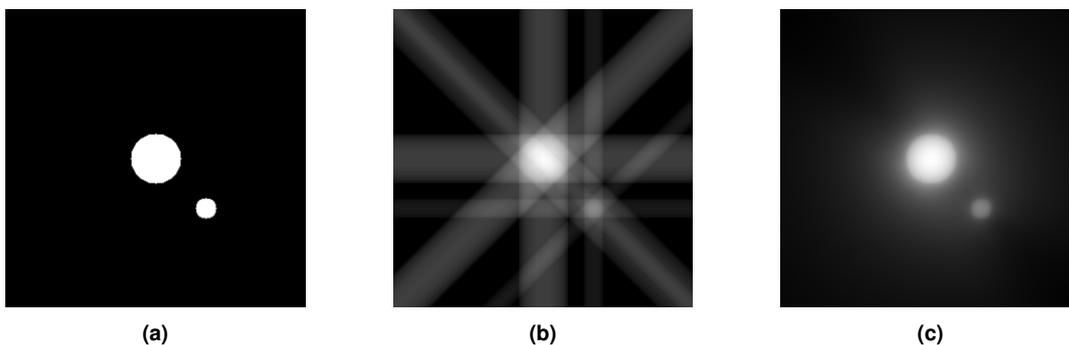


Figure 3. Backprojection image generation, in (a) the original image, in (b) the summ of projections from 0° to 179° in steps of 45° and in (c) the summ with steps of 1° .

Here the final image was constructed by rotating the lesion image in inverted grayscale in 180° with 1° steps, which covers the entire image. This method provided the elimination of noise present in the images and emphasized the lesion region. Thus,

the final image is a representation of the initial image, but with less definition. After obtaining the backprojection image, Otsu's threshold binarization was applied. Figure 4a shows the backprojection image from Figure 2b and Figure 4b shows the result of the binarization process by Otsu's thresholding.

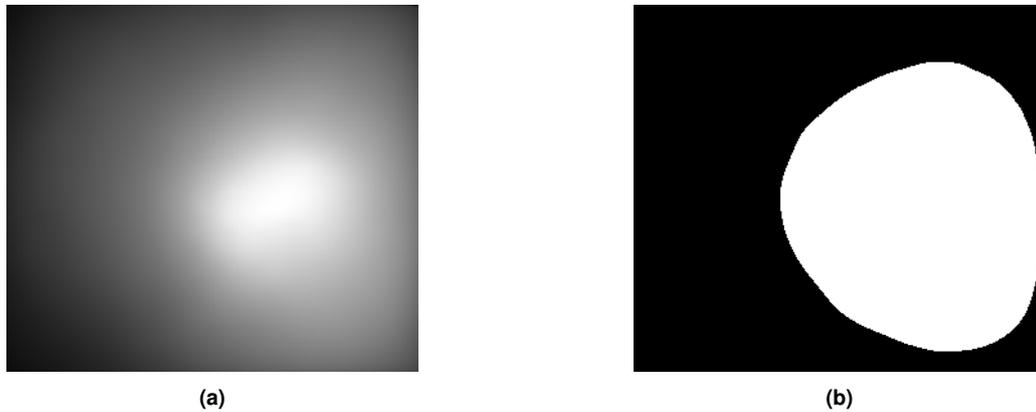


Figure 4. Segmentation of the backprojection image, in (a) the backprojection image and in (b) the binarized image by Otsu's thresholding.

2.2.3. Segmentation by Otsu's Thresholding on Grayscale Subtracted and Filtered Images

The third method consisted of subtracting the backprojection image from the normalized grayscale image, Figure 5a. The binarization by the Otsu's threshold was done in the subtracted image, however, only in the region defined by the mask generated by the second method, which can be seen in Figures 5b and 5c. The final segmentation definition was done by searching for the smallest contour of the Otsu's binarized image that contained the center of the lesion. The center of the lesion was defined as the binarized white region with a fixed threshold of 250 in the backprojection image, Figure 6a. The contour image and the final segmentation, generated by the third method, can be seen in Figures 6b and 6c.

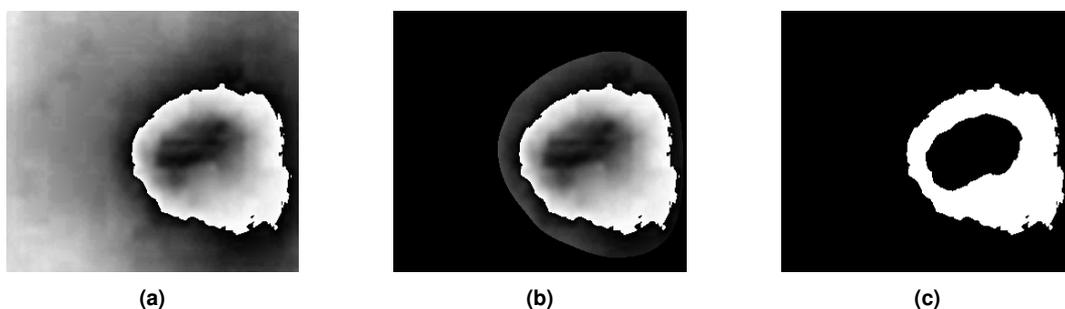


Figure 5. Segmentation on the masked image, in (a) the result of subtracting the backprojection image from the grayscale normalized, in (b) the application of the backprojection mask on the subtracted image and in (c) the binarization of the masked image by Otsu's thresholding.

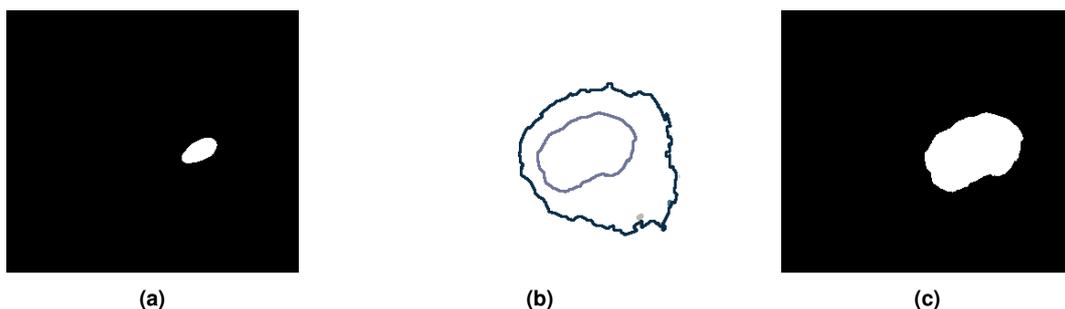


Figure 6. Definition of the lesion contour, in (a) the segmentation with fixed threshold on the backprojection image, in (b) the contours found on binarized image after subtraction and masking processes and in (c) the segmentation result from the 3rd method.

2.3. Evaluation Metrics

The methods were evaluated according to the classification of each pixel on the image. This classification was compared to the ground truth and the pixel was considered: true positive (TP), if it has correctly classified a lesion; true negative (TN), if it has correctly classified the mucosa; false positive (FP), if it has incorrectly classified the mucosa as lesion; and false negative (FN), if it has incorrectly classified the lesion as mucosa.

To compare the results obtained by each method, the metrics of accuracy, precision, recall and f1-score or dice-score, were calculated pixel by pixel against the ground truth. With special attention to the f1-score, which is a widely used metric to assess the quality in the segmentation of images [Tanriver et al. 2021], as it indicates the overlap ratio of the segmentation with the ground truth.

3. Results

After applying the three methods in all images, the segmentation performance was evaluated. In Table 1, the averages of accuracy, precision, recall and f1-score for the 21 images are presented for each method.

Table 1. Average results comparison based on the metrics from tested methods

Metric	1st Method	2nd Method	3rd Method
Accuracy	66.1%	66.5%	84.9%
Precision	41.3%	38.6%	70.1%
Recall	92.1%	97.1%	75.3%
F1-score	48.6%	48.5%	60.8%

It can be seen that the third method presented better results in accuracy, precision and f1-score, behind only in the evaluation of recall, where the method that stood out the most was the second. Comparing the results obtained in each of the metrics and observing a few example cases, some advantages and disadvantages of the methods can be highlighted.

3.1. Lesion Location

From the three methods used, it is noticeable that the second and third methods performed better in locating the lesion in the image. The first method, Otsu's thresholding on the grayscale inverted image, had difficulty in discriminating between lesion and other structures, because it only takes into consideration the intensity of the pixels. This causes artifacts such as saliva bubbles and other reflections to be characterized as lesions. The second method, thanks to the use of the backprojection image, was able to locate the lesion region on the image, helping to eliminate some artifacts such as reflections, but there was a difficulty in segmenting the borders. This is because the image generated by the backprojection technique accentuates the lesion region, but has little definition regarding the contours. The third method brought an improvement in the definition of contours, eliminating most of the pixels erroneously classified as lesion by the second method.

An example of these problems can be seen in Figure 7, which brings an image with artifacts such as saliva bubbles and the result of segmentation by the three methods described. In this case, the second method showed a single region containing the lesion, while the first method showed multiple regions. Meanwhile, the third method brings an improvement over the second method by being able to reduce the amount of pixels incorrectly classified as lesion.

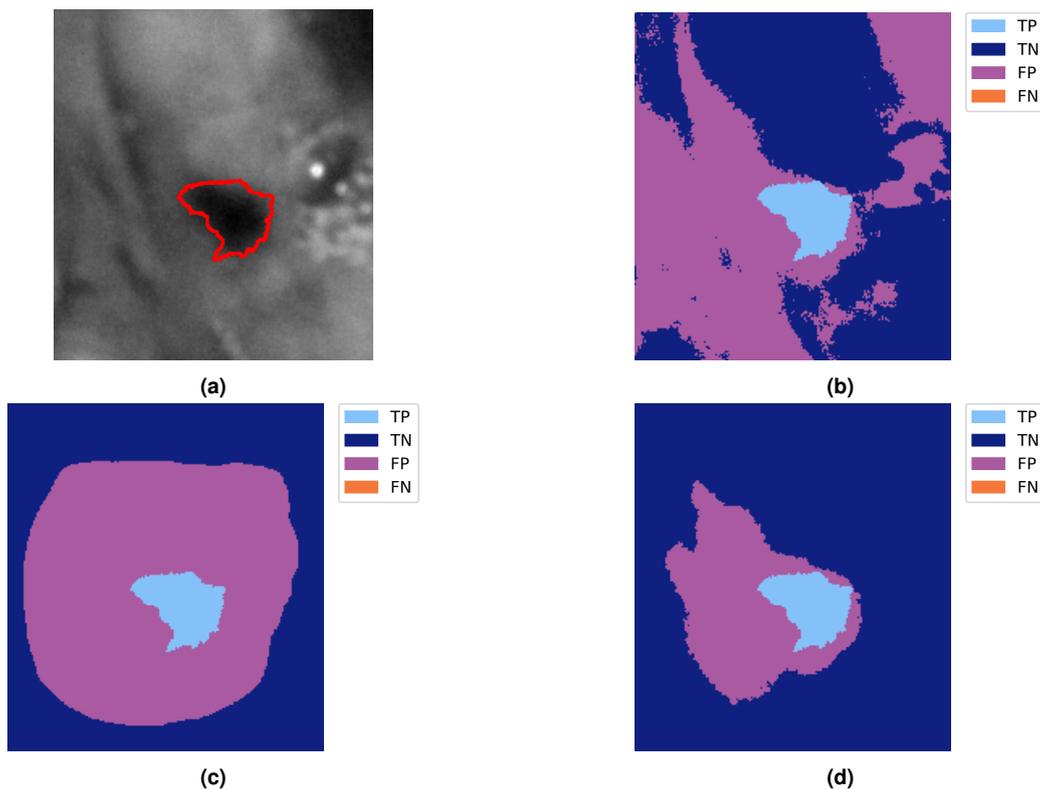


Figure 7. Methods comparison by the pixel's classifications for macule 18, in (a) the grayscale normalized image from lesion with defined ground truth, in (b) the result from 1st method segmentation, in (c) the result from 2nd method segmentation and in (d) the result from 3rd method segmentation.

In the images where the lesion is highlighted with a good contrast and there are no other artifacts, such as bubbles or reflections, the first method obtained good results, as

can be seen in Figure 8. The histogram of this image showed good similarity to a bimodal model, with good separation between classes and a more uniform distribution between dark and light pixels.

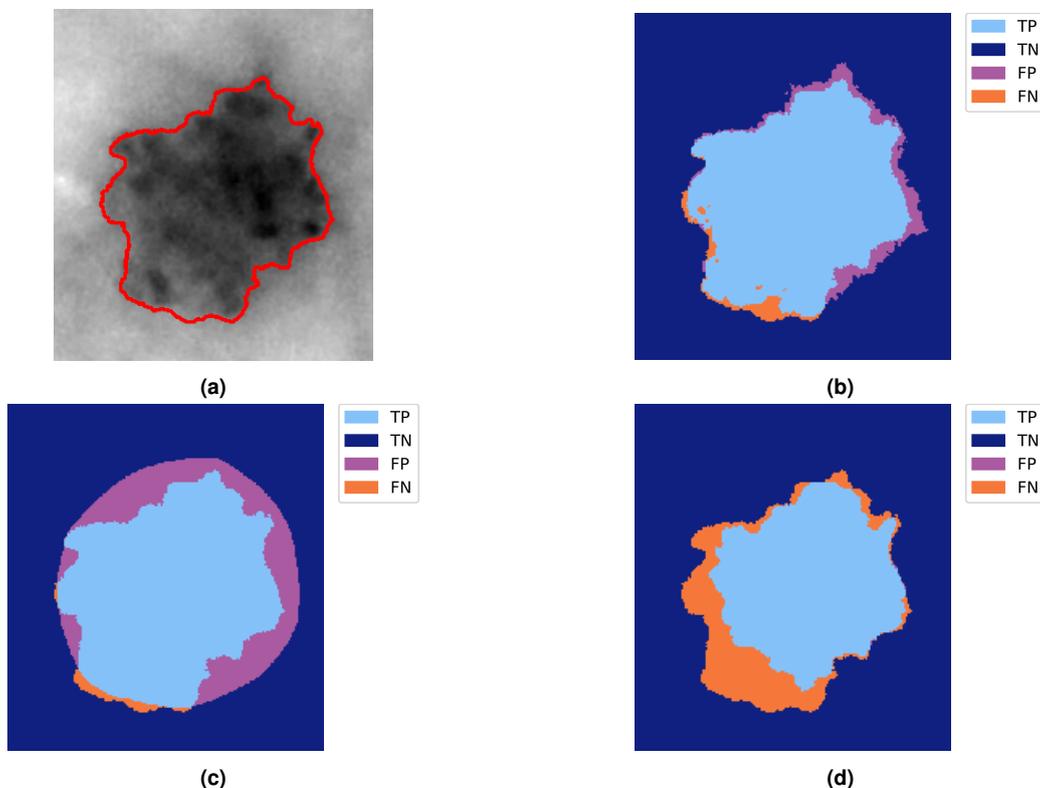


Figure 8. Methods comparison by the pixel's classifications for macule 3, in (a) the grayscale normalized image from lesion with defined ground truth, in (b) the result from 1st method segmentation, in (c) the result from 2nd method segmentation and in (d) the result from 3rd method segmentation.

In this case, another point to be observed, is that the third method presented a larger amount of false negative pixels, classifying as lesion a smaller region than defined by the other methods. This shows an important characteristic, which is the proportion that the lesion occupies in the image.

3.2. Lesion and Background Ratio

While the lesion in Figure 7a occupies a small portion of the image, the lesion in Figure 8a occupies a large portion. In Figure 7c, the second method was able to encompass 100% of the lesion within the area given as lesion, but ends up identifying a large region of healthy mucosa as the lesion. The third method in this case was able to better isolate the lesion, Figure 7d, decreasing the region of false positives. But in Figure 8c, the second method classified some pixels as false negatives, meaning that the method classified them as healthy when they were actually lesion. In such cases, the third method, Figure 8d, inherits the difficulties of the second method for lesions that occupy a large portion of the image, since it uses the segmentation of the second method as a mask to locate the lesion.

With the application of the third method, the segmentation edges improve, being more similar to those of the lesion, but the amount of false negatives increases, especially

in cases where the lesions occupy a large area of the image. This shows a difficulty of the second and third methods in adjusting to this proportion between lesion and image background.

3.3. Image Luminance and Contrast

The images in the database present the most different patterns of luminance, contrast and sharpness, making it difficult for a method to work with similar results for all cases. We chose to use the Otsu-based global thresholding, because it is a well known and consolidated method, but it is based on the fact that the image histogram must be bimodal, which in many cases it does not happen.

In this work the image was inverted for some procedures, which is a form of histogram equalization that inverts the gray levels, but does not change the division of the classes by the Otsu's threshold. In addition, normalization of the images to gray levels from 0 to 255 was performed, not significantly altering the shape of the histogram wrapper. In some cases, this procedure improved the contrast between background and lesion, but in some images with artifacts this was not possible. Tests were made with Otsu's thresholding segmentation on pre-processed images. On the inverted grayscale images the gamma function was applied. The gamma function, Equation 1, is a function used for histogram equalization.

$$I[l, c] = 255 \times \left(\frac{I[l, c]}{255} \right)^\gamma \quad (1)$$

Where I is the intensity value of the pixel in the l row and c column of the image. If the γ coefficient is greater than 1, it darkens the image by making dark pixels darker. If the coefficient is less than 1, it has the opposite effect, emphasizing the lighter pixels.

Figure 9, presents the results of the segmentation by the first method after the equalization of the histogram by the gamma function with a coefficient equal to 5. It is possible to compare the images with those segmented by the first method without equalization, Figures 7b and 8b.

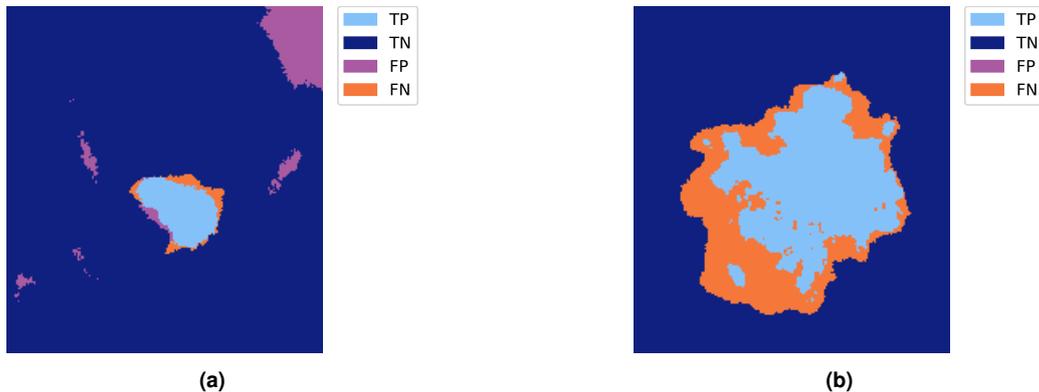


Figure 9. Otsu's segmentation after equalization, in (a) the result of the process applied to Figure 7a and in (b) the result of the process applied to Figure 8a.

It is perceptible that the lesion in Figure 9a showed a better result with equalization, but Figure 9b got a worse result. This is connected to the average luminance

of the images and the presence of artifacts. The presence of artifacts such as reflections worsens the linear contrast adjustment and, in such cases, nonlinear functions can get better results, making segmentation easier. This means that the application of histogram equalization techniques may improve the results, but it should be done based on previous information about the image, such as mean luminance or histogram density distribution.

4. Discussion

It is noteworthy that the databases used in this research and in others, cited here, were not the same, mainly because there are not much options of open database with these type of lesions. Besides this fact, segmentation techniques must be as generalized as possible, in a way not to lose performance due to different data insertions. Thus, the methods presented here should be further tested against broader image databases.

However, looking at the obtained results and comparing them to the related works, it may be seen that the third method had results that resemble the results of some works, even using less complex techniques. Anantharaman et al. (2018) developed a Mask R-CNN with a ResNet-101 base that obtained a dice score of 77.4% for labial herpes and 71.4% for aphthous ulcers. The third method presented here obtained a dice score of 60.8% for images of oral macules and spots, a value considerably close, even without the use of deep machine learning. The dice score found by Tanriver et al. (2021) reached a value of 92.6% with a U-Net with EfficientNet-b7. Despite this, the paper suggests using a YOLOv5l model that achieved an average accuracy of 64.4% for selecting a rectangular area. This is due to the high processing and time required for segmentation by the other networks tested.

Other works have used texture as a way to segment and classify lesions, however, the objects of the study were different. Santos et al. (2012), developed a system for semi-automatic segmentation of skin lesions and Thomas et al. (2013) developed a work for semi-automatic segmentation of lesions caused by carcinoma, which are mostly characterized by ulcerated or verrucous lesions. These lesions are characterized by having different colors and textures, which do not occur in oral macules or spots. Pigmented lesions, as macules and spots, in many cases present borders that gradually change color until reaching the area of healthy mucosa, which may explain the false negative or false positive areas at the margins of the lesion. In this case, an arbitrary binary definition for gradual transition is challenging. The task becomes even more difficult if one considers the fact that macules can have different tones and the healthy mucosa can also exhibit different shades, varying for each individual. But considering that the cited works used semi-automatic segmentations, this could be a future alternative for color segmentation in macular lesions.

Observing the results of the segmentation by the third method, one can see that the variation of the proportion between background and lesion in an image has an influence on the number of false positives found. This demonstrates that the method can be improved in order to adapt to different lesion proportions in the image. One proposal would be to evaluate other forms of thresholding than Otsu's method. This is because it is based on the clear density of the probability that a pixel belongs to a certain class, which makes it difficult for the method to fit small lesion regions. The images also have different qualities, with varying illumination, sharpness, and contrast, which makes it difficult to generalize

the segmentation process. In the future, pre-processing focused on contrast enhancement based on the image histogram distribution may bring benefits to the segmentation.

The segmentation by the third method was able to eliminate noise and much of the region that did not compose the lesion. Thus, it presents itself as a pre-processing method for a lesion classification step that can then use an artificial intelligence network. This allows the elimination of points that may be confusing to the network, improving the classification results. Finally, it is noted that there is a link between works that perform segmentation of oral and dermatological lesions, which could be another application for the third method presented, bringing a low-processing alternative to existing solutions so far.

5. Conclusion

With the developed work, it was possible to evaluate three methods for macule and oral spot segmentation. The method that best performed, considering all the images, was the third one, whose process involves the creation of a backprojection image of the lesion image in grayscale, inverted and normalized, followed by the binarization by Otsu's thresholding technique in two stages. The only metric where the third method did not perform the best was recall, where the second method performed better. Still, the third method came close to the results of works developed with deep machine learning, which is interesting since the method developed here has a lower complexity. Furthermore, as it is a method that does not use machine learning, it brings the advantage of not requiring previous training, since it is difficult to find databases with images of oral macules of good quality. For the future, these methods should be tested with other images and different databases.

References

- Anantharaman, R., Velazquez, M., and Lee, Y. (2018). Utilizing mask r-cnn for detection and segmentation of oral diseases. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2197–2204.
- dos Santos, E. S., de M S Veras, R., R T Aires, K., M B F Portela, H., Braz Junior, G., Santos, J. D., and Tavares, J. M. R. (2022). Semi-automatic segmentation of skin lesions based on superpixels and hybrid texture information. *Medical Image Analysis*, 77:102363.
- Gomes, R. F. T., Schmith, J., Figueiredo, R. M. d., Freitas, S. A., Machado, G. N., Romanini, J., and Carrard, V. C. (2023a). Use of artificial intelligence in the classification of elementary oral lesions from clinical images. *International Journal of Environmental Research and Public Health*, 20(5).
- Gomes, R. F. T., Schuch, L. F., Martins, M. D., Honório, E. F., de Figueiredo, R. M., Schmith, J., Machado, G. N., and Carrard, V. C. (2023b). Use of deep neural networks in the detection and automated classification of lesions using clinical images in ophthalmology, dermatology, and oral medicine—a systematic review. *Journal of Digital Imaging*.
- Liu, Y.-J., Yu, C.-C., Yu, M.-J., and He, Y. (2016). Manifold slic: A fast method to compute content-sensitive superpixels. pages 651–659.

- Mahmood, H., Shaban, M., Rajpoot, N., and Khurram, S. A. (2021). Artificial intelligence-based methods in head and neck cancer diagnosis: an overview. *British Journal of Cancer*, 124:1934–1940.
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66.
- Sarode, G., Maniyar, N., Sarode, S. C., Jafer, M., Patil, S., and Awan, K. H. (2020). Epidemiologic aspects of oral cancer. *Disease-a-Month*, 66(12):100988. Oral Health Special Issue.
- Song, B., Li, S., Sunny, S., Gurushanth, K., Mendonca, P., Mukhia, N., Patrick, S., Peterson, T., Gurudath, S., Raghavan, S., Tsusennaro, I., Leivon, S. T., Kolar, T., Shetty, V., Bushan, V., Ramesh, R., Pillai, V., Wilder-Smith, P., Suresh, A., Kuriakose, M. A., Birur, P., and Liang, R. (2022). Exploring uncertainty measures in convolutional neural network for semantic segmentation of oral cancer images. *Journal of Biomedical Optics*, 27(11):115001.
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 71(3):209–249.
- Tanriver, G., Soluk Tekkesin, M., and Ergen, O. (2021). Automated detection and classification of oral lesions using deep learning to detect oral potentially malignant disorders. *Cancers*, 13(11).
- TelessáudeRS (2022). Quem somos.
- Thomas, B., Kumar, V., and Saini, S. (2013). Texture analysis based segmentation and classification of oral cancer lesions in color images using ann. In *2013 IEEE International Conference on Signal Processing, Computing and Control (ISPC)*, pages 1–5.
- Welikala, R. A., Remagnino, P., Lim, J. H., Chan, C. S., Rajendran, S., Kallarakkal, T. G., Zain, R. B., Jayasinghe, R. D., Rimal, J., Kerr, A. R., Amtha, R., Patil, K., Tilakaratne, W. M., Gibson, J., Cheong, S. C., and Barman, S. A. (2020). Automated detection and classification of oral lesions using deep learning for early detection of oral cancer. *IEEE Access*, 8:132677–132693.
- World Medical Association (2013). World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. *JAMA*, 310(20):2191–2194.