Critical analysis of the performance of the YOLO architecture in the detection of oral lesions from clinical images

Gustavo Goetz Ribeiro[1], **Jean Schmith**[1,4,5], **Rita F. T. Gomes**[2],
Giovanna Nunes Machado[1], **Vinicius C. Carrard**[2,3], **Rodrigo Marques de Figueiredo**[1]

[1]Polytechnic School, Unisinos University
São Leopoldo, RS, Brazil

[2]Departament of Oral Pathology, Faculdade de Odontologia
Federal University of Rio Grande do Sul (UFRGS)
Porto Alegre, RS, Brazil

[3]TelessaúdeRS, Federal University of Rio Grande do Sul (UFRGS)
Porto Alegre, RS, Brazil

[4]SENAI Innovation Institute for Sensor Systems (ISI-SIM)
São Leopoldo - RS - Brazil

[5]Competence Center on Digital Agriculture (EMBRAPII / SENAI-RS)
São Leopoldo - RS - Brazil

gustavogoetz@gmail.com, j.schmith@gmail.com, ritafabgomes@yahoo.com.br

giovannanm@edu.unisinos.br, vccarrard@gmail.com, marquesf@unisinos.br

Abstract. *The objective of this work was to train a convolutional neural network for the detection of oral cavity lesions in heterogeneous clinical images, using YOLOv5, as well as the critical analysis of its effectiveness. The database had four categories of elementary oral lesions, without stipulating protocols for obtaining the images, such as distance, angle, and illumination. YOLO showed the best performance in detecting vesicular/blister lesions in both models analyzed: YOLOv5m mAP@50 was 86.8% and in the YOLOv5x model it was 80.7%, followed by papule/nodule in both tests. Images containing only one lesion showed better performance. We considered the quality of the detections obtained to be satisfactory in the majority of images despite using a small dataset for this evaluation.*

1. Introduction

In dentistry, the recognition of oral lesions is fundamental in professional practice, regardless of the specialty. Several studies show that clinical dentists perceive themselves as poorly prepared to recognize and diagnose stomatological lesions, especially malignant lesions at an early stage. This finding points to the need to think about strategies that reverse this situation and invest in public policies that support actions aimed at reducing morbidity and mortality from oral cancer [Nam et al. 2018].

Oral cancer has an enormous impact on the quality of life of a patient. Its treatment is aggressive and is associated with an approximate survival of 60% in patients in general, and 39% in patients diagnosed at an advanced stage [Warin et al. 2022]. The anatomical structures of the oral cavity present a wide chromatic and textural variation, which can occur under physiological or pathological conditions, making the evaluation process challenging due to the wide number of possibilities that should be considered as a differential diagnosis for a professional with little experience [Güneri and Epstein 2014].

In recent years, the potential of technological resources that involve artificial intelligence (AI) to assist in the early diagnosis process has been discussed. Incorporation of these resources is based on the possibility of automated recognition of characteristics suggestive of malignancy from the analysis of clinical (photographic) images [Gomes et al. 2023a]. In recent years, AI methods have been shown to be useful in detecting, classifying, and segmenting objects from medical images [Gomes et al. 2023b][Kelsch et al. 2023].

[Lin et al. 2021] found that most oral lesions in the database are relatively small, i.e. they do not occupy most of the image. This means that the captured image may have many irrelevant backgrounds. The size of the lesion in the image may vary with different distances between the camera and the lesion or with the use of different cameras and different focal lengths. This causes a great variability in the performance of the AI system. To obtain better results in automated classification of clinical images lesions, image pre-processing, such as cropping the area of the lesion, becomes essential for the good performance of convolutional neural network algorithms, especially when the area of interest is embedded in a complex image [Gomes et al. 2023a]. Unwanted information from the image must be removed to provide only the information needed for decision making. Several different approaches to this process have been used and often a combination of several techniques has been used to improve accuracy [Ilhan et al. 2021].

[Tanriver et al. 2021] used semantic segmentation to delimit lesions in images. It is important to emphasize that semantic segmentation delimits but does not differentiate what is a lesion from what is not a lesion, making it necessary to combine two techniques – segmentation and object detection. To avoid irrelevant information in the image, [Fu et al. 2020] used a detection network that took an oral photograph as input and generated a boundary box that located the suspected lesion. The lesion area was cropped and the selected area was used to feed a classification network.

The YOLO – You Only Look Once architecture was developed as a method to detect single-stage objects. YOLO detects objects as a regression problem directly from image pixels to bounding boxes and their respective classes. YOLOv5 is the fifth published version of YOLO in which it was developed by Ultralytics. YOLOv5 has pre-trained models, including YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x [Jocher et al. 2023]. Since the literature presented promising results on YOLOv5 [Tanriver et al. 2021] considering oral lesions, more of these versions should be explored. Therefore, the objective of this work was to train a YOLOv5 for the detection of oral cavity lesions in heterogeneous clinical images. Several versions of YOLOv5 were tested and a detailed analysis of results is provided by a stomatology specialist. For this work, it was not limited to the number of lesions or the centralization of the lesion, but to a marking that considers all occurrences of the lesion in the image, as well as the critical analysis of its effectiveness.

	Papule/Nodule	**Vesicle/Blister**	**Ulcer**	**Plaque**	**Total**
Images per category	268	73	305	391	1037
Removing the test images	258	63	295	381	997
Data augmentation	382	378	389	385	1534
Labels per class	551	600	771	749	2671
Images for training	301	300	308	305	1214
Images for validation	81	78	81	80	320

Table 1. Detailed description of the dataset numbers.

2. Methodology

This research protocol follows the principles of the Declaration of Helsinki, has been evaluated and approved by the local Human Research Ethics Committee (GPPG No. 2019 0746). The image selection period was from November 1, 2021 to January 28, 2022, by an expert researcher with 8 years of experience (third author of this work). A senior specialist (the fifth author of this work) supervised and assisted the other researchers in case of doubt. The images were also labeled by the same specialists. The database consists of images containing four categories of described elementary oral lesions, namely: papule / nodule, vesicles / blister, ulcer and plaque. Elemental lesions are a clinical classification used for soft tissue lesions. This classification is visual and clinical, without the need for complementary tests, thus serving as the gold standard for validation [Gomes et al. 2023a].

For [Mortazavi et al. 2019], the category of papule / nodule is described as a superficial, elevated, and solid lesion. The vesicle / blister is a superficial, elevated lesion with fluid content inside. Ulcer is characterized by the loss of continuity of the epithelium in which the center of the lesion is initially red, which can turn white / grayish. Ulcers can be deep or superficial. The plaque is characterized by a superficial and slightly elevated lesion.

The images were obtained using professional or semi-professional cameras or smartphone cameras, without stipulating any protocol for obtaining the images, such as distance, angle, and lighting. The images were captured by several dental surgeons, without any type of previous calibration. From a visual inspection, images that did not have good focus or luminosity were discarded. The division of the dataset is presented in Table 1.

Ten images from each category of lesion were selected to be used in a qualitative analysis of the results, totaling 40 test images. The selection was for convenience, where we selected examples from different positions and locations of the oral cavity. To balance the amount of data in all categories of lesions, the data augmentation technique was used. The rotation, vertical and horizontal mirroring of the images were used as data augmentation techniques, obtaining a dataset of 1534 images. The reason for choosing ”lesion” as the only class for the detection model was first the low number of images from each lesion, and second is a strategy to verify how general the model could be. With data augmentation, in addition to increasing the number of images, the number of labels also increased since many images have more than one lesion marked. Table 1 shows the total number of images used for training and validation and the gains obtained after the data augmentation process.

Several images had more than one lesion, and a label was made for each lesion present in the image, i.e. they have multiple lesions labeled in each of the images. The open source software LabelImg was used for labeling, as this software allows exporting the file with the coordinates in the TXT format supported in YOLOv5. After the labbeling process, the ulcer images have 771 labels and the plaque images 759 labels, the nodule / plaque images had 551 labels, and the vesicle / blister images had 600 labels.

Experiments were carried out using the YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x models. For network training, the Google Colab platform was used with a Nvidia Tesla T4 GPU (Graphics Processing Unit) and with Python programming libraries related to machine learning. It should be noted that for training, all lesions were defined as a single class, that is, YOLOv5 only detects if there is a lesion in the tissue or not.

To evaluate the performance of the object detector, precision, recall, average precision (AP), and intersection over union (IOU) were used as evaluation metrics. For the validation of the detector performance for the four categories of lesions individually, 320 lesion images were separated. The number of images of each category used in this evaluation is described in Table 1.

The results of the test images were sent to an oral medicine specialist with 8 years of experience to perform a visual and qualitative analysis of the obtained results. Initially, 3 levels of difficulty were determined for detection on images, namely: difficult level (presents lesions that do not have well-defined limits); medium level (presents a more diffuse, irregular lesion); easy level (presents a well-defined lesion). In addition, the professional evaluated the quality of the detections generated by the YOLOv5m and YOLOv5x models, highlighting whether the result was satisfactory, partially satisfactory, or unsatisfactory in the marking generated in the image.

3. Results

Initially, we tested four versions of YOLOv5 to analyze which of them have a better performance. Next, we present the results of the two best models for each category of elemental oral lesion analyzed in this study. After several tests, the best results were achieved with 300 epochs, batch size of 16 and all images with size of 640 x 640 pixels. We tested the models YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x and the best results were achieved with YOLOv5m and YOLOv5x with a learning rate of 0.01 and a final learning rate of 0.1.

All 320 images in the validation database had at least one lesion. When checking the detections in this database for the YOLOv5m model, it was observed that in 151 images there were no detections. For the YOLOv5x model, there were no detections in 62 images. Table 2 presents the results for each category of lesion when using the YOLOv5m and YOLOv5x models.

YOLO showed better performance in the detection of vesicle / blister lesions in the two models analyzed, in which in the YOLOv5m model the mAP@50 was 86.8% and in the YOLOv5x model 80.7%. The second best performance was in the detection of papule/nodule in both tests. For the YOLOv5m model, the resulting mAP@50 was 44.8% and for the YOLOv5x model, 48.7%. Regarding the others categories of lesions, ulcer and

Model	Metric	Papule / Nodule	Vesicle / Blister	Ulcer	Plaque
YOLOv5m	Precision	0.834	**0.961**	0.356	0.337
	Recall	0.340	**0.756**	0.339	**0.395**
	mAP@50	0.448	**0.868**	**0.361**	**0.370**
	mAP@50:95	**0.316**	**0.521**	**0.161**	**0.158**
YOLOv5x	Precision	**0.861**	0.878	**0.584**	**0.499**
	Recall	**0.418**	**0.756**	**0.346**	0.362
	mAP@50	**0.487**	0.807	**0.361**	0.347
	mAP@50:95	0.300	0.480	0.142	0.140

Table 2. Results by lesion with YOLOv5m and YOLOv5x considering the validation dataset.

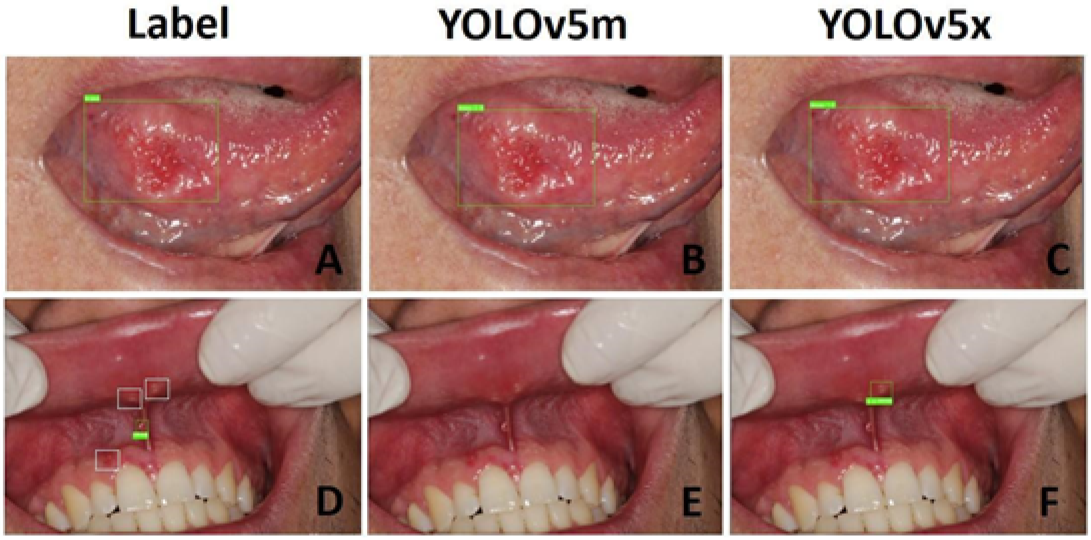

Figure 1. Demonstrates the pattern of labeling (A) and detections (B and C) expected and in D, E and F, the gaps observed between labeling and detection. The white rectangles in D represent images of small lesions underestimated at the labeling stage, partially detected by the YOLOv5x model (F).

plaque, the results were very similar in the two tests, however in the YOLOv5m test, the plaque lesion obtained the third best result with mAP@50 of 37% and the ulcer the worst performance with mAP@50 36.1%. Regarding YOLOv5x validation, the ulcer lesion had the third best performance with mAP@50 36.1% and plaque with mAP@50 34.7%. Figure 1 demonstrates the pattern of labeling (A) and detections (B and C) expected, and in D, E and F, the gaps observed between labeling and detection.

We performed a visual analysis of the detections obtained in the YOLOv5m and YOLOv5x models for the 320 images of the validation dataset, in order to account for the number of true positives (PV), false positives (FP) and false negatives (FN). The 320 images were marked for the two models analyzed using a confidence threshold 25%. Only detections with a confidence value greater than or equal to 25% were generated. Observing Figure 2, we can verify inconsistencies where, although the label was not considered at certain locations, when analyzing the delimitations of the models, they were compatible with the lesions and were considered false positives, or even where there were

Model	Metric	Papule / Nodule	Vesicle / Blister	Ulcer	Plaque
YOLOv5m	VP	41	53	42	33
	FP	4	2	1	1
	FN	89	27	174	115
	Precision	0.911	0.964	0.977	0.971
	Recall	0.315	0.663	0.194	0.223
YOLOv5x	VP	63	59	91	79
	FP	7	16	6	10
	FN	79	14	129	74
	Precision	0.900	0.787	0.938	0.888
	Recall	0.444	0.808	0.414	0.516

Table 3. Performance of YOLOv5m and YOLOv5x markings considering the validation dataset.

		n	%
Difficulty level	Easy	21	52.5 %
	Medium	5	12.5 %
	Hard	14	35.0 %
Lesions in the image	Unique	27	67.5 %
	Multiple	13	32.5 %
		YOLOv5m	**YOLOv5x**
Performance	Complete marking	25 (62.5 %)	27 (67.5 %)
	Incomplete marking	9 (22.5 %)	9 (22.5 %)
	No marking	6 (15.0 %)	4 (10.0 %)

Table 4. Performance of YOLOv5m and YOLOv5x markings considering the characteristics of the images of the test dataset.

three labels, the models performed a single delimitation that included all lesions of the labels. Taking into account the label classification, the results are presented in Table 3.

Based on a qualitative analysis of the test dataset, the images of the lesions were classified first by a specialist according to the level of difficulty of visualization and the definition of contours and limits. In total, 35% of the lesions in the images were considered difficult to identify. In 67.5% of the images, only one lesion could be identified. The delimitations were considered adequate in 62.5% of the YOLOv5m tests, and 67.5% of the YOLOv5x tests, however, considering the performance of the models in the delimitation of the lesions of interest, the YOLOv5x proved to be more effective, as can be seen in Table 4.

Table 5 presents the performance analysis according to the levels of difficulty in identifying and delimiting the lesions contained in the images, as well as the performance in the case of single or multiple lesions in the images, confirming that the images considered easy and those containing only one lesion presented satisfactory performance in the two models considered.

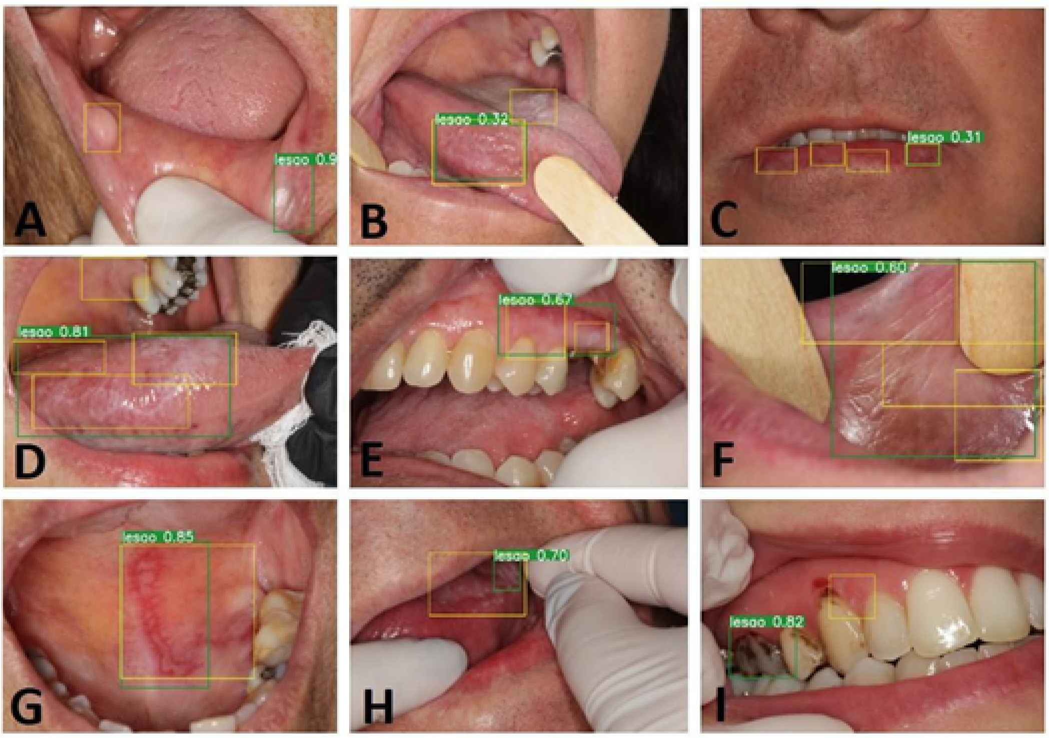

Figure 2. The figure shows the assigned labels in yellow and the detections of the models in green. In A, an erroneous detection of the model. In B and C, there were false negatives when demarcating only one lesion, when there were labels for multiple lesions. In D, E and F, the model detected only one demarcation encompassing individual labels. In G and H, the demarcations did not cover the lesion in its entirety, however it encompassed significant areas. In I, there was a false positive when demarcating a tooth element.

		Good in both	Good for YOLOv5m	Good for YOLOv5x	Partial in both	Poor in both
Difficulty level	Easy	5	2	1	4	2
	Medium	2	1	2	0	0
	Hard	18	0	0	2	1
Lesions in the image	Unique	21	1	1	1	3
	Multiple	4	0	4	5	0

Table 5. Qualitative performance of lesion markings of the test dataset.

4. Discussion

The performance of AI in image processing depends greatly on the method used and the type of data analyzed. Data preparation requires an expert to select the most useful resources for training and task performance [Aljuaid and Anwar 2022], where all relevant information can be considered and where system objectives are clearly defined so that decision making based on data or algorithms works perfectly [Hassani et al. 2020]. The possible interpretations and functionalities of the method that is intended to be used must be considered.

For [Fu et al. 2020], [Tanriver et al. 2021] and [Alzubaidi et al. 2021] a critical point refers to the heterogeneity of oral disease lesions. The variability in data quality is highlighted, such as lighting, zoom, angle, sharpness, and resolution. [Figueroa et al. 2022] warns about the influence of irrelevant information in the general context of the image. Irrelevant information may attract the attention of the algorithm, despite being related to the lesion. The proposed algorithm fails to distinguish several visually confusing cases, however, for it to have practical applicability, it is necessary to seek strategies to overcome these challenges, as it is difficult to strictly follow research protocols in clinical practice, with multiple users.

According to [Welikala et al. 2020], the performance of the algorithm tends to improve with increasing data. This study agrees with the complex characteristic of the oral cavity image and the heterogeneous nature of the lesions, since, in addition to the complex background, each image can contain one or more lesions. The results showed that the YOLOv5 models had low recall values, with the YOLOv5m model presenting a recall of only 31.2% and the YOLOv5x model a recall of 42.3%, which indicates a limitation in the ability to detect all relevant cases, indicating the appearance of false negatives.

The images were classified according to the level of complexity in visualizing the contours and limits of the lesions to understand the reasons why the experiment had a high rate of false negatives. In total, 35% of the images, the lesions were considered difficult to identify (Table 4). In 67.5% of the images, only one lesion could be identified. The delimitations were considered adequate in 62.5% of YOLOv5m and 67.5% of the YOLOv5x tests. [Lin et al. 2021] found that the majority of the lesions in their database are relatively small, that is, they do not occupy the majority of the image. This means that the captured image may contain a lot of irrelevant information. The size of the image lesion can vary due to several factors such as: different distances between the camera and the lesion; use of different cameras or the original size of the lesion itself. We also observed this issue in our work, and these variations cause great variability in model performance.

Typically, in healthcare, biological data tends to be unbalanced; the volume of non-disease images is more common than images of certain diseases. It should be noted that undesirable results could be produced when training an IA model using imbalanced data. The study by [Song et al. 2021] found the influence of unbalanced datasets on the performance of AI-based oral cancer classification, where minority classes had significant classification errors. Some authors consider the widely accepted data augmentation technique to balance the data and increase the robustness of the model [Fu et al. 2020]. On the other hand, according to [Welikala et al. 2020], the need for an approach that increases the size of the data set is a limitation.

The results presented here show high recall in the vesicle/bubble category. The performance of this class stands out compared to other classes. We can corroborate that this phenomenon refers to the fact that this class predominantly presents a single lesion per image, and may also be an influence of the data augmentation process, since it was the class that presented a smaller sample number. The other classes also underwent data augmentation, however, the proportion of synthetic data generated was lower. Although it is a widely accepted technique for balancing data, it can influence model performance [Gomes et al. 2024] [Maccagnan et al. 2023].

To define whether a detection is correct or not, the neural network uses the IOU between the predicted detection and the label defined in the image. We analyze the detections used in the validation to evaluate the results obtained. Figure 1 (A) presents an example of an image with a single lesion in which both models (B and C) obtained satisfactory performance. Figure 1 D presents a negative result when the image had only one lesion. In this example, YOLOv5m (E) did not detect any lesion while YOLOv5x (F) incorrectly detected a lesion, generating a false positive. Analyzing this image, we observed an influence of data preparation. Images that presented more than one lesion were labeled and presented separately to the algorithm. Both models made partial detections, pointing to the need to rethink the way training data is prepared, dedicating efforts to proposing strategies to improve this performance. An approach strategy for this issue was proposed by [Welikala et al. 2020] and repeated by [Rajendran et al. 2023], to obtain more data, 3 to 7 independent annotators labeled the images, each annotation being considered new data. However, this strategy may lead to repeated information.

The evaluation metrics described in the previous sections were generated from the results of the validation dataset. Lesions that were not detected were considered false negatives by the algorithm, which affects the Recall value obtained and consequently the mAP@50 value. Perhaps the success rate would be higher if, in the labeling stage, marking was allowed that involved all small lesions, with a single marking window, highlighting the importance of understanding how the model was trained and what errors were generated. The performance analysis stage goes far beyond numbers and metrics, one need to understand why a mistake was done and try to correct the error, enhancing positive results. [Warin et al. 2022] analyzed the data generated by defining lesion boundaries labeled by 3 maxillofacial surgeons, due to differences in manual segmentation between surgeons, the ground truth used in AI training, validation and testing was the largest intersection area between all surgeons' notes.

From a practical point of view, to truly define the effectiveness of lesion detection, it is necessary to consider the objective for which the detection is proposed. Considering that we only want to say whether there is a lesion or not, if the model makes a broader or more restricted marking, as long as it encompasses the lesion, there are no reservations. Furthermore, if the model is intended to evaluate the evolution of the lesion, in this case the need for precise limits is justified. Examples of this can be seen in Figure 2 (D, E and F). The model detections obtained from the 320 validation images, the YOLOv5m model obtained 177 detections of which 169 (95.48%) correspond to correct detections (true positives) and 8 (4.52%) to incorrect detections (false positives). The YOLOv5x model obtained 331 detections, of which 292 (88.22%) corresponded to correct detections and

39 (11.78%) to incorrect detections. In this way, false positives and false negatives can be questioned and reconsidered in a qualitative analysis to guide the algorithm's response that can be accepted as correct.

The dataset included 65.5% images with a single lesion and 32.5% images with multiple lesions. In images with multiple lesions, YOLOv5 needed to make multiple markings to achieve acceptable scores, however, as shown in Figure 2, alternative markings did not necessarily mean an error, highlighting that it is possible to obtain better performance in the models, with the adoption of alternatives to correct this pattern of interpretation. Among images with a single lesion, we observed better performance in both tests, but although the images had a single label, that is, with a single lesion, YOLOv5 identified underestimated lesions, as illustrated in Figures 1 - D and F.

In general, the accuracy values performed well, with the models achieving an accuracy of 78.9% for YOLOv5m and 62.1% for YOLOv5x. These values indicate that YOLOv5 has a good ability to detect true positives and few false positives, despite the clinical images showing a complex environment. This concept seeks to demonstrate how AI and humans can complement each other and co-exist in a mutually beneficial way. AI can promote the increase of human capabilities [Hassani et al. 2020], but it does not replace the human component [Warin et al. 2022]. As a limitation of the study, we can consider the preparation of data that must predict complex imaging situations with a lot of irrelevant but eye-catching information and the possibility that an image encompasses several lesions.

5. Conclusion

Through a qualitative evaluation of the performance of a neural network using the YOLOv5 architecture by a specialist in oral medicine, we considered the quality of the detections obtained to be satisfactory in the majority of images, despite having used a small dataset for this evaluation. When evaluating the detections of the validation dataset with the defined labels, a direct relationship was observed between the quality of the label and the performance in the metrics generated by YOLO. For future work, it is suggested to expand the marking options accepted as correct. Another suggestion is to continue the work and implement a second stage to classify the injuries obtained according to the category of each lesion. Moreover, more recent versions of YOLO should be tested in these kinds of images. Finally, other detection model approaches such as DETR (Detection Transformer), ConvNeXt, ViTDet, and Deformable DETR should also be tested [Flügge et al. 2023, Warin and Suebnukarn 2024].

Acknowledgment

This work has partially funded and supported by the Competence Center on Digital Agriculture, with financial resources from PPI HardwareBR of the MCTI grant number 056/2023, signed with EMBRAPII.

References

Aljuaid, A. and Anwar, M. (2022). Survey of supervised learning for medical image processing. *SN Computer Science*, 3(4):292.

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., and Farhan, L. (2021). Review of deep learning: concepts, cnn architectures, challenges, applications, future directions. *Journal of big Data*, 8:1–74.

Figueroa, K. C., Song, B., Sunny, S., Li, S., Gurushanth, K., Mendonca, P., Mukhia, N., Patrick, S., Gurudath, S., Raghavan, S., et al. (2022). Interpretable deep learning approach for oral cancer classification using guided attention inference network. *Journal of biomedical optics*, 27(1):015001–015001.

Flügge, T., Gaudin, R., Sabatakakis, A., Tröltzsch, D., Heiland, M., van Nistelrooij, N., and Vinayahalingam, S. (2023). Detection of oral squamous cell carcinoma in clinical photographs using a vision transformer. *Scientific Reports*, 13(1):2296.

Fu, Q., Chen, Y., Li, Z., Jing, Q., Hu, C., Liu, H., Bao, J., Hong, Y., Shi, T., Li, K., et al. (2020). A deep learning algorithm for detection of oral cavity squamous cell carcinoma from photographic images: A retrospective study. *EClinicalMedicine*, 27.

Gomes, R. F. T., Schmith, J., de Figueiredo, R. M., Freitas, S. A., Machado, G. N., Romanini, J., Almeida, J. D., Pereira, C. T., de Almeida Rodrigues, J., and Carrard, V. C. (2024). Convolutional neural network misclassification analysis in oral lesions: an error evaluation criterion by image characteristics. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, 137(3):243–252.

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):3894.

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*, 36(3):1060–1070.

Güneri, P. and Epstein, J. B. (2014). Late stage diagnosis of oral cancer: components and possible solutions. *Oral oncology*, 50(12):1131–1136.

Hassani, H., Silva, E. S., Unger, S., TajMazinani, M., and Mac Feely, S. (2020). Artificial intelligence (ai) or intelligence augmentation (ia): what is the future? *Ai*, 1(2):8.

Ilhan, B., Guneri, P., and Wilder-Smith, P. (2021). The contribution of artificial intelligence to reducing the diagnostic delay in oral cancer. *Oral oncology*, 116:105254.

Jocher, G. et al. (2023). Yolov5 by ultralytics. 2020. `https://github.com/ultralytics/YOLOv5`.

Kelsch, C. R., Schmith, J., Gomes, R. F., Carrard, V. C., and de Figueiredo, R. M. (2023). Image processing methods for oral macules and spots segmentation. In *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, pages 256–267. SBC.

Lin, H., Chen, H., Weng, L., Shao, J., and Lin, J. (2021). Automatic detection of oral cancer in smartphone-based images using deep learning for early diagnosis. *Journal of Biomedical Optics*, 26(8):086007–086007.

Maccagnan, G. C., Schmith, J., Santos, M., and de Figueiredo, R. M. (2023). Toolbox for vessel x-ray angiography images simulation. In *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, pages 59–70. SBC.

Mortazavi, H., Baharvand, M., Dalaie, K., Faraji, M., Khalighi, H., and Behnaz, M. (2019). Oral lesion description: a mini review. *International Journal of Medical Reviews*, 6(3):81–87.

Nam, Y., Kim, H.-G., and Kho, H.-S. (2018). Differential diagnosis of jaw pain using informatics technology. *Journal of Oral Rehabilitation*, 45(8):581–588.

Rajendran, S., Lim, J. H., Yogalingam, K., Kallarakkal, T. G., Zain, R. B., Jayasinghe, R. D., Rimal, J., Kerr, A. R., Amtha, R., Patil, K., et al. (2023). Image collection and annotation platforms to establish a multi-source database of oral lesions. *Oral Diseases*, 29(5):2230–2238.

Song, B., Li, S., Sunny, S., Gurushanth, K., Mendonca, P., Mukhia, N., Patrick, S., Gurudath, S., Raghavan, S., Tsusennaro, I., et al. (2021). Classification of imbalanced oral cancer image data from high-risk population. *Journal of biomedical optics*, 26(10):105001–105001.

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):2766.

Warin, K., Limprasert, W., Suebnukarn, S., Jinaporntham, S., Jantana, P., and Vicharueang, S. (2022). Ai-based analysis of oral lesions using novel deep convolutional neural networks for early detection of oral cancer. *Plos one*, 17(8):e0273508.

Warin, K. and Suebnukarn, S. (2024). Deep learning in oral cancer-a systematic review. *BMC Oral Health*, 24(1):212.

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., et al. (2020). Automated detection and classification of oral lesions using deep learning for early detection of oral cancer. *Ieee Access*, 8:132677–132693.