Enhanced Detection and Segmentation of Diabetic Retinopathy Lesions Using a Region of Interest Strategy

Marcelo Dias[1], **Carlos Santos**[2,3], **Alejandro Pereira**[1], **Marilton Aguiar**[1], **Daniel Welfer**[3]

[1] Postgraduate Program in Computing –
Federal University of Pelotas (UFPEL) – Pelotas, RS – Brazil

[2]Federal Institute of Education –
Science and Technology Farroupilha (IFFar) – Alegrete, RS – Brazil

[3]Postgraduate Program in Computer Science –
Federal University of Santa Maria (UFSM) – Santa Maria, RS – Brazil

{adspereira, marcelo.sdias, marilton}@inf.ufpel.edu.br,
carlos.santos@iffarroupilha.edu.br,
daniel.welfer@ufsm.br

Abstract. *Diabetic retinopathy is a common microvascular complication of diabetes that can lead to vision loss. Experts diagnose it by identifying retinal lesions, including hard exudates, soft exudates, hemorrhages, and microaneurysms, the earliest indicators. These lesions often appear near the macula, making this region crucial for early diagnosis. This study proposes a YOLOv12-based approach for detecting and segmenting retinal lesions, focusing on the macula. We trained two models: one for region extraction using the Indian Diabetic Retinopathy Dataset and another for lesion detection with the Dataset for Diabetic Retinopathy. The model achieved a mean Average Precision of 0.4530 on the validation set and 0.3020 on the testing set for lesion segmentation.*

1. Introduction

Diabetic retinopathy (DR) is a complication of diabetes that affects the small blood vessels in the retina, resulting from prolonged high blood glucose levels [Alyoubi et al. 2021]. It impacts one in three people with diabetes and is a leading cause of vision loss among adults [of Ophthalmology 2017]. While there is currently no cure for DR, the risk of blindness can be reduced through early detection of retinal lesions. This can be achieved using fundus images taken by retinal cameras [Kumar and Ramaswamy Karthikeyan 2021].

The diagnosis of DR depends on the analysis of fundus images, where various lesions – such as microaneurysms (MA), hemorrhages (HE), hard exudates (EX), and soft exudates (SE) – are identified. Microaneurysms appear as small, round lesions and are the first clinical sign of DR, playing a crucial role in monitoring the progression of the disease [Murugan and Roy 2022]. However, detecting microaneurysms remains challenging due to their shape variations and similarity to other retinal lesions.

This work proposes a deep learning-based pipeline to enhance the detection and segmentation of retinal lesions. The approach incorporates pre-processing techniques

concentrating on regions of interest within the fundus, specifically the macula and fovea. A key contribution of this study is the introduction of a training strategy that targets a specific region of the eye and integrates image pre-processing, resulting in more efficient lesion detection in DR. This method has shown promising results, particularly in detecting MAs, which are among the most challenging types of lesions to identify.

2. Related Work

The study conducted by [Munuera-Gifre et al. 2020] investigates which regions of the eye are most vulnerable to retinal lesions. The findings indicate a clear pattern among patients with DR, showing that MA, HE, and EX occur more frequently in the left eye compared to the right. Additionally, lesion density is significantly higher in the central retina of the left eye, particularly in the macular region. The study suggests that microaneurysms initially develop in the temporal cells near the macula, which are critical for forming lesions.

According to [Ometto et al. 2017], initial lesions typically present as small red dots, indicating the presence of MA and/or HE that develop in the macular area, specifically to the temporal side of the fovea. The study reports that DR lesions first appear in this region, suggesting that their presence is an early indicator of diabetic maculopathy.

The study by [Santos et al. 2023] introduces an approach for segmenting retinal lesion instances using Detectron2 [Wu et al. 2019]. This method includes several pre-processing steps, such as partially removing the black background through cropping, dividing the image into blocks (tiling), and applying data augmentation techniques that feature color and geometric transformations. The approach was evaluated on the Dataset for Diabetic Retinopathy (DDR) [Li et al. 2019] and the Indian Diabetic Retinopathy Image Dataset (IDRiD) [Porwal et al. 2018]. It demonstrated promising results in detecting and segmenting MA, EX, SE, and HE. The model achieved a mean average precision (mAP) of 0.2941 during validation and 0.1735 during testing on the DDR dataset.

A study by [Pereira et al. 2023] utilized the single-stage detector YOLOR [Wang et al. 2021] to identify retinal lesions in fundus images using the DDR and IDRiD datasets. To enhance the model's performance, the researchers implemented various pre-processing techniques, including cropping and Slicing-Aided Inference (SAHI), which functions like a "sliding window" over the original image. This method resulted in significant mAP values: 0.3280 for EX, 0.2030 for SE, 0.2190 for HE, and 0.1400 for another category, culminating in an overall mAP of 0.2225 on the test set.

The studies conducted by [Munuera-Gifre et al. 2020] and [Ometto et al. 2017] examine the spatial distribution of retinal lesions in DR, highlighting the macula as a crucial region. The model can extract relevant features more effectively by concentrating on this area, where lesions are most commonly found, thus improving detection accuracy. Similarly, [Santos et al. 2023] and [Pereira et al. 2023] achieved promising results by segmenting images into smaller blocks to enhance feature extraction. While [Santos et al. 2023] utilized a 2×2 division, [Pereira et al. 2023] implemented a 4×4 block strategy. However, despite these advantages, such methods significantly increase the number of images used during training, including less relevant regions, ultimately raising computational costs.

3. YOLOv12

YOLOv12 [Tian et al. 2025] represents the latest advancement in the YOLO series, designed as an attention-centric real-time object detector. It maintains the speed of previous CNN-based models while leveraging the enhanced performance provided by attention mechanisms [Tian et al. 2025]. The architecture of YOLOv12 consists of three main components: 1) *Backbone* incorporates area attention (A2) to capture global dependencies while reducing computational complexity; 2) *Neck* utilizes residual efficient layer aggregation networks (R-ELAN) to optimize feature fusion; and, 3) *Head* enhances final predictions through an optimized attention-based structure [Tian et al. 2025]. YOLOv12 introduces several significant advancements, including FlashAttention for efficient memory access, a simplified hierarchical design that eliminates unnecessary positional encoding, and modifications to the MLP ratio to balance computational cost. Across all model scales, YOLOv12 consistently outperforms previous versions and competing real-time detectors, resulting in improved accuracy and computational efficiency [Tian et al. 2025].

4. Materials and methods

The model was trained on a system equipped with an Intel Core i7-10700F processor (2.90 GHz, 16 threads), 16 GB of RAM, and a GeForce RTX 3060 graphics card with 12 GB of VRAM. This study utilizes the YOLOv12 model [Tian et al. 2025] for lesion detection and segmentation in fundus images, while the YOLOv11 model [Jocher and Qiu 2024] is used for detecting the macula region. This approach facilitates the extraction of regions of interest (ROIs).

4.1. Datasets

We utilized two publicly available datasets for our experiments: DDR and IDRID. The IDRID dataset comprises 516 images that include annotations for the optic disc and the fovea region, which are critical for generating regions of interest. Meanwhile, the DDR dataset contains 757 images with pixel-level segmentations and bounding box annotations for lesions classified as EX, HE, SE, and MA.

4.2. Performance Metrics

To evaluate object detection models, assessing their predictive performance on a dataset is crucial. In this study, we utilized the Average Precision (AP) metric, which is widely used in object detection tasks [Santos et al. 2021]. AP is calculated as the area under the Precision-Recall (PR) curve, where precision indicates the proportion of correct predictions, and recall measures the detection rate for all positive class instances. Additionally, we employed the mean Average Precision (mAP) metric, which represents the average AP across all classes in the dataset [Santos et al. 2022]. Equation 1 outlines the calculation of mAP.

$$mAP = \frac{1}{n} \sum_{i=1}^{n} AP_i \tag{1}$$

4.3. Pre-processing

This section outlines the two image pre-processing pipelines utilized in this work. The first pipeline focuses on training the model to detect the optic disc and macula. The

second pipeline describes the pre-processing steps to train the model for detecting and segmenting DR-associated retinal lesions.

4.3.1. Pre-processing for Optic disc and macula detection

In this stage, we based our approach on studies by [Santos et al. 2022] and [Murugan and Roy 2022] for processing data to extract the region of interest. Initially, we applied a cropping technique to partially remove the black background in the images. Additionally, we examined different color spaces to improve feature extraction. We found that the green channel in the RGB color space provided better contrast between blood vessels and the macula, which supported its use in the image analysis.

After completing the previous step, we applied a median filter to remove impulsive noise from the images. Next, we utilized the Contrast Limited Adaptive Histogram Equalization (CLAHE) technique to enhance image quality. Finally, a bilateral filter was used to smooth the images while preserving important details. This pre-processing step aims to maximize the visibility of the macula, which appears as the darkest region in the images. Figure 1 illustrates the pre-processing pipeline.

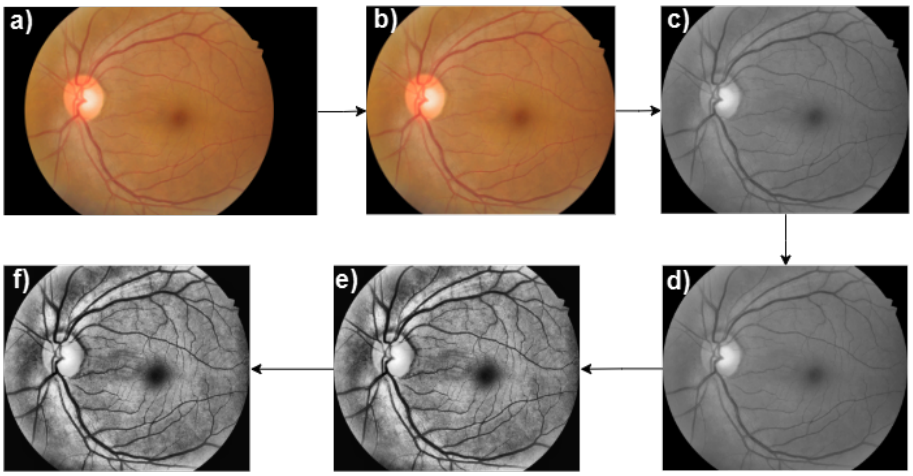

Figure 1. Pre-processing pipeline for fundus images: a) Original image; b) Crop background partially; c) Extract the green channel; d) Apply median filter; e) Use CLAHE; f) Apply bilateral filter. Source: Author.

4.3.2. Pre-processing for Retinal Lesions Detection

In our approach to image pre-processing for retinal lesion detection, we utilized the following pipeline: (1) median filtering, (2) CLAHE, (3) partial background removal (cropping), and (4) region of interest extraction. First, we applied a median filter to smooth the image and reduce potential noise. Next, CLAHE enhanced image contrast by converting the image to the LAB color space and performing histogram equalization on the luminance channel. This method preserved the color information while improving local contrast without excessively amplifying noise. After these adjustments, the contrast-enhanced image returned to its original color space. Figure 2 illustrates the complete pre-processing pipeline we adopted.

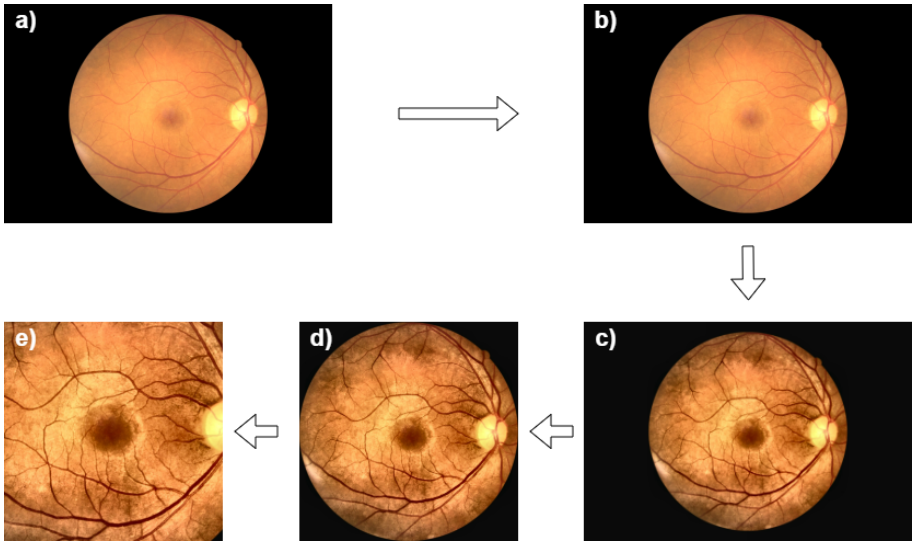

Figure 2. Image pre-processing for training the retinal lesion detection model: a) Original image; b) Median filtering; c) CLAHE; d) Cropping; e) Region of interest extraction. Source: Author.

4.4. Extraction of the region of interest

This study stage focuses on detecting the macula and the optic disc. While the primary emphasis is on the macula, the coordinates of the optic disc can serve as a reference point for estimating the macula's position in cases where direct detection is not possible. The IDRID dataset includes the coordinates of the fovea, which represents the central area of the macula. These coordinates allow for a reliable estimation of the surrounding macular region.

We chose YOLOv11 [Jocher and Qiu 2024] for ROI extraction because of its robustness, speed, and accuracy. As the task of detecting the macula and the optic disc focuses on locating well-defined anatomical structures, unlike lesion segmentation which involves identifying multiple small and heterogeneous pathological patterns, we opted for a lighter and faster model than YOLOv12. The IDRID dataset was split into training, validation, and test subsets in the proportions of 50 percent, 20 percent, and 30 percent respectively, resulting in 258 images for training, 103 for validation, and 155 for testing.

The model was fine-tuned on the validation set to optimize hyperparameters for detecting the macula and optic disc. Training was conducted using a batch size of 32 over 1,000 epochs, with early stopping implemented after 100 epochs to prevent overfitting. Images were resized to 320×320 pixels, and the pre-trained YOLOv11x model was utilized along with the AdamW optimizer and a learning rate of 0.001. To further mitigate overfitting, a dropout rate of 10% was applied. The final hyperparameters, fine-tuned based on validation results, are presented in Table 1. Additionally, several data augmentation techniques were employed to enhance the model's generalization and robustness. These techniques included random rotations of up to 15 degrees, translations, horizontal and vertical flips, resized cropping, and the mosaic technique, which combines four training images into a single composition.

Following the completion of the training phase, we assessed the model, which demonstrated promising results in detecting both the macula and the optic disc. The model

Table 1. Hyperparameters used for training the model for macula detection on the IDRID dataset.

Hyperparameters	Value
Learning Rate	0.001
Batch Size	32
Epochs	1000
Patience	100
Input Size	320
Dropout Rate	10%

achieved mean Average Precision at IoU 0.50 (mAP@50) values of 0.99 for the optic disc and 0.95 for the macula on the test set. The results for both datasets are presented in Table 2.

Table 2. Results obtained for optic disc and macula detection on the validation and test sets of the IDRID dataset using the mAP@50 metric.

Dataset	Optic Disk	Macula	mAP
Validation	0.99	0.96	0.97
Test	0.99	0.95	0.97

With the model trained and validated, the next step was to apply it to the DDR dataset to generate ROIs. The DDR dataset consists of 757 fundus images, and the model achieved an accuracy of 96.40% in detecting the macula, demonstrating good generalization from its training on the IDRID dataset. To address the 27 images in which the macula was not detected, we implemented a strategy based on the structural characteristics of the eye. According to [Elloumi and Kachouri 2023], the distance between the macula and the optic disc is approximately 2.5 times the diameter of the optic disc. This understanding allowed us to infer its position. As a result, we successfully identified the region of interest in 756 images, with only one image remaining undetected and subsequently discarded due to its low quality.

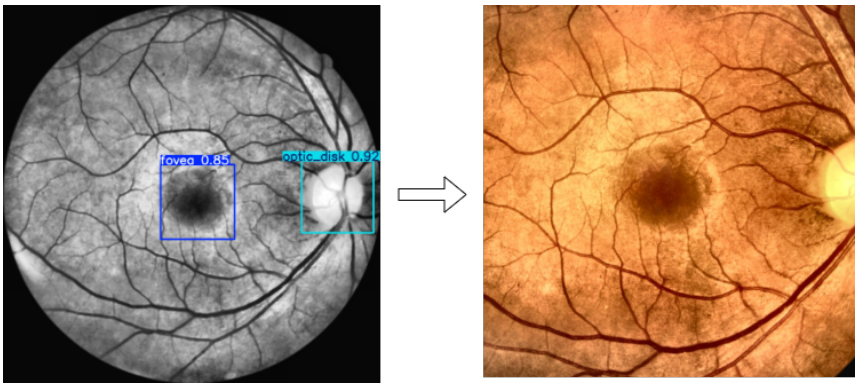

Figure 3. Extraction of the region of interest by using the central coordinates of the macula and applying them to the original image. Source: Author.

Figure 3 illustrates an example of the model's predictions for macula and optic disc detection within the DDR dataset, showcasing how the region of interest was extracted using the bounding box coordinates generated by the prediction.

4.5. Training YOLOv12 for detecting retinal lesions and performing instance segmentation using ROIs

After completing the image processing pipeline, we trained the computational model for retinal lesion detection. We divided the DDR dataset into training, validation, and test sets, using proportions of 50%, 20%, and 30%, respectively. The model, YOLOv12 [Tian et al. 2025], was trained for 400 epochs, with early stopping implemented at 50. We used a batch size of 4 and resized the images to 800×800 pixels. To reduce the risk of overfitting, we applied a dropout rate of 10%. The hyperparameters utilized for model training are detailed in Table 3.

Table 3. Hyperparameters used for training the model to detect instance segmentation of retinal lesions (EX, HE, SE, and MA) in the DDR Dataset.

Hyperparameters	Value
Optimizer	AdamW
Learning Rate	0.001
Momentum	0.9
Batch Size	4
Epochs	400
Patience	50
Input Size	800
Dropout Rate	10%

We utilized various data augmentation techniques during training to enhance the model's robustness and improve its generalization ability. These techniques included horizontal and vertical flips, rotations, translations, and resized cropping. In addition to these fundamental transformations, we implemented more advanced methods such as Mosaic, which merges four different images into a single composition; Copy-Paste, which extracts objects from one image and inserts them into another; and Mixup, which blends two images along with their labels in a weighted manner.

4.6. Evaluation

During model training, we utilized ROIs to focus on areas with a higher likelihood of containing lesions. However, this strategy limited our access to the broader context of the full image, which could impact the model's evaluation. To facilitate a more thorough analysis and to account for all present lesions, we implemented a tiling strategy. This approach divides the image into smaller blocks, enabling detections across various regions. Consequently, this method ensures that the evaluation process considers all lesions, leading to a more comprehensive analysis. Figure 4 illustrates how tiling is applied to the original image.

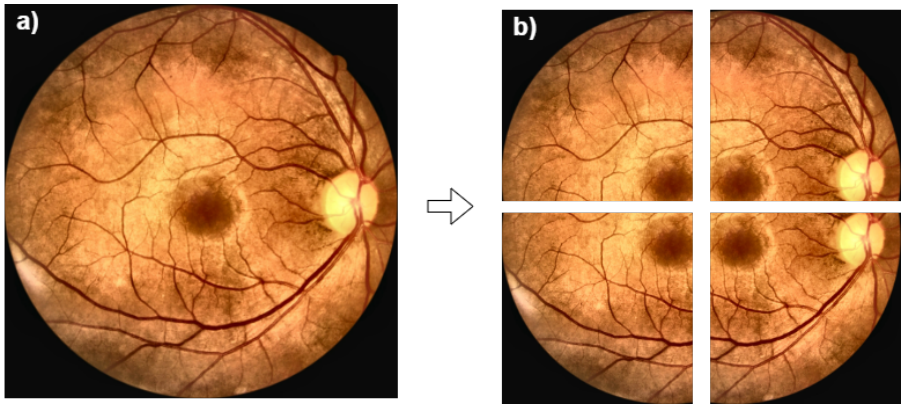

Figure 4. Presentation of image processing techniques to enhance model assessment: a) Original image; b) Image with tiling applied. Source: Author.

5. Results

We compared our proposed model with state-of-the-art single-stage object detection models. In our experiments, we evaluated the following models: YOLOv8 [Jocher et al. 2023], YOLOv9 [Wang et al. 2024b], YOLOv10 [Wang et al. 2024a], YOLOv11 [Jocher and Qiu 2024], and YOLOv12 [Tian et al. 2025]. We tested all models using pre-trained weights and selected their deeper versions, which have a higher number of parameters. Additionally, we compared our results with the work of [Santos et al. 2023], who employed Detectron2 [Wu et al. 2019], and [Pereira et al. 2023], who used YOLOR [Wang et al. 2021].

Tables 4 and 5 show the results for object detection and instance segmentation of EX, HE, SE, and MA lesions using the mAP@50 metric on the validation and test sets. The results include both bounding boxes (bbox) and segmentation masks (mask), with the highlighted values indicating the best performance for each lesion type.

Table 4. Results obtained for detection (BBox) and instance segmentation (Mask) using the proposed method, with AP and mAP@50 on the DDR validation Dataset.

Model		AP				mAP
		EX	HE	SE	MA	
YOLOv8 [Jocher et al. 2023]	BBox	0.3220	0.4310	0.5420	0.1760	0.3680
	Mask	0.2310	0.2960	**0.5230**	0.0766	0.2820
YOLOv9 [Wang et al. 2024b]	BBox	0.3420	0.4360	**0.5470**	0.1800	0.3740
	Mask	0.2580	0.3360	0.5040	0.0692	0.2920
YOLOv10 [Wang et al. 2024a]	BBox	0.3400	0.4340	0.5130	0.1810	0.3680
	Mask	–	–	–	–	–
YOLOv11 [Jocher and Qiu 2024]	BBox	0.3510	0.4450	0.5420	0.1690	0.3770
	Mask	0.2410	0.3380	0.5060	0.0627	0.2870
YOLOv12 [Tian et al. 2025]	BBox	0.3300	0.4400	0.5360	0.1790	0.3710
	Mask	0.2170	0.3320	0.5030	0.0445	0.2740
Detectron2 [Santos et al. 2023]	BBox	0.2941	0.3520	0.3508	0.1644	0.2903
	Mask	0.3012	0.3211	0.3505	0.2036	0.2941
YOLOR-CSP [Pereira et al. 2023]	BBox	0.4090	0.3870	0.4660	0.2610	0.3808
	Mask	–	–	–	–	–
Proposed approach	BBox	**0.4280**	**0.4720**	0.4990	**0.3480**	**0.4370**
	Mask	**0.4390**	**0.4680**	0.5070	**0.3980**	**0.4530**

Table 5. Results obtained for detection (BBox) and instance segmentation (Mask) using the proposed method, with AP and mAP@50 on the DDR test Dataset.

Model		AP				mAP
		EX	HE	SE	MA	
YOLOv8 [Jocher et al. 2023]	BBox	0.2880	0.2890	0.2320	0.0851	0.2240
	Mask	0.2020	0.1860	0.2180	0.0435	0.1620
YOLOv9 [Wang et al. 2024b]	BBox	0.2860	0.3120	**0.2620**	0.0980	0.2400
	Mask	0.2170	0.2230	0.2260	0.0469	0.1780
YOLOv10 [Wang et al. 2024a]	BBox	0.2890	0.3060	0.2270	0.0956	0.2290
	Mask	–	–	–	–	–
YOLOv11 [Jocher and Qiu 2024]	BBox	0.3080	0.3090	0.2420	0.1000	0.2400
	Mask	0.2110	0.1980	0.2160	0.0314	0.1640
YOLOv12 [Tian et al. 2025]	BBox	0.2910	0.3200	0.2270	0.0837	0.2300
	Mask	0.1920	0.2040	0.2130	0.0268	0.1590
Detectron2 [Santos et al. 2023]	BBox	0.2515	0.1577	0.1548	0.1042	0.1670
	Mask	0.2687	0.1388	0.1589	0.1274	0.1735
YOLOR-CSP [Pereira et al. 2023]	BBox	0.3280	0.2190	0.2030	0.1400	0.2225
	Mask	–	–	–	–	–
Proposed approach	BBox	**0.3960**	**0.3680**	0.2280	**0.1770**	**0.2920**
	Mask	**0.4080**	**0.3570**	**0.2350**	**0.2080**	**0.3020**

Figure 5 illustrates the model's training progress. There is a steady decrease in classification loss (cls_loss), bounding box regression loss (box_loss), segmentation loss (seg_loss) and distribution focal loss (dfl_loss), along with an increase in precision, recall, mAP@50, and mAP@50-95 metrics. These trends indicate consistent learning and improved detection of lesions. Furthermore, the close alignment of the training and validation loss curves suggests that the model generalizes well without clear signs of overfitting.

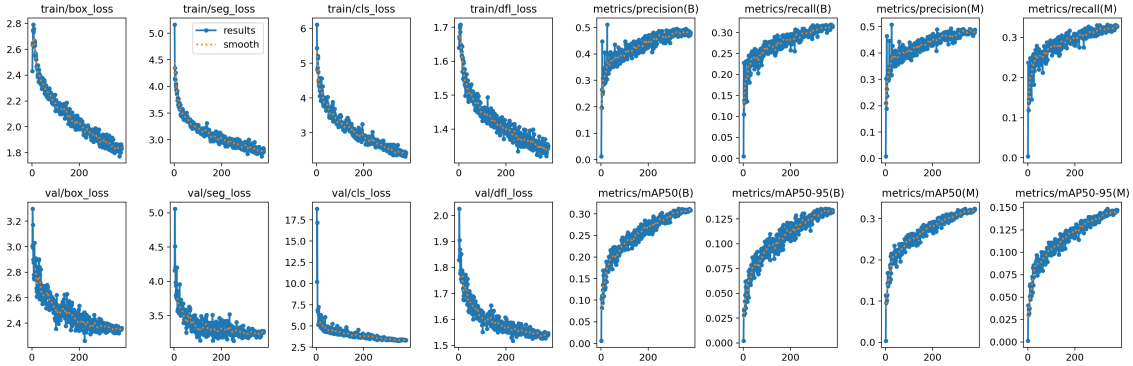

Figure 5. Presentation of metrics: box_loss, seg_loss, cls_loss, dfl_loss, Precision, and Recall generated during the training of the proposed model. Source: Author.

Upon analyzing Tables 4 and 5, we can see that the proposed method has shown promising results in the detection and segmentation of retinal lesions. In the DDR test set, the AP values for segmentation were 0.4080 for EX lesions, 0.3570 for HE lesions, 0.2350 for SE lesions, and 0.2080 for MA lesions. In terms of detection, the results were 0.3960, 0.3680, 0.2280, and 0.1770, respectively. In the validation set, the results were even more impressive, with AP values of 0.4390 for segmentation of EX lesions, 0.4680 for HE lesions, 0.5070 for SE lesions, and 0.3980 for MA lesions. The detection results

in the validation set were also noteworthy, yielding AP values of 0.4280 for EX lesions, 0.4720 for HE lesions, 0.4990 for SE lesions, and 0.3480 for MA lesions.

When comparing performance on the validation and test sets, we found that exploring ROIs led to improved outcomes. Our approach outperformed the competing models in detecting and segmenting EX, HE, and MA lesions. Notably, it showed significant advancements in detecting and segmenting MA lesions, which are among the most challenging to identify. The best competing model achieved an AP of 0.2036 for segmentation and 0.2610 for detection on the validation set, and scores of 0.1274 and 0.1400, respectively, on the test set. In contrast, our proposed approach reached an AP of 0.3980 and 0.3480 in the validation set, and 0.2080 and 0.1770 in the test set, demonstrating substantial improvements. Figure 6 presents an image from the test set illustrating the model's predictions.

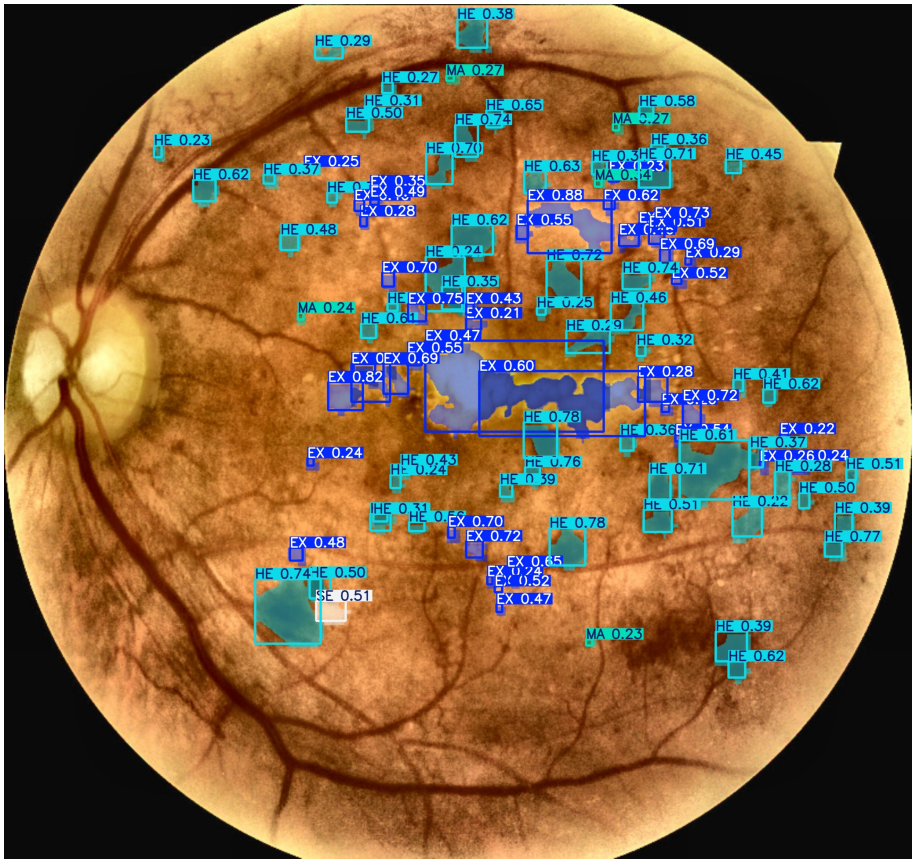

Figure 6. Example of the model's predictions for retinal lesion detection and segmentation in an image from the DDR test set. The lesion types are EX (hard exudates), HE (hemorrhages), SE (soft exudates), and MA (microaneurysms). Source: Author.

Despite the promising results, the model demonstrated lower performance in detecting and segmenting SE lesions, achieving satisfactory results only in segmentation on the test set. This outcome can be attributed to two main factors: (1) the limited number of SE examples in the dataset, which hampers effective learning, and (2) the lower occurrence of SE in the macular region, reducing its exposure during training. Additionally, despite some improvements, detecting MA remains challenging due to their small size.

6. Conclusion

This study presents a pre-processing pipeline that targets the macular region to extract areas of interest. It eliminates regions with a low likelihood of containing lesions, thereby increasing the model's focus on the most relevant areas. Furthermore, pre-processing techniques were employed to improve the detection and segmentation of retinal lesions.

We utilized two datasets in our study: the IDRID dataset for training the model that is responsible for extracting regions of interest, and the DDR dataset for evaluating the model's performance in detecting and segmenting lesions classified as EX, HE, SE, and MA. In our experiments, the model demonstrated promising results. The performance metrics were as follows: 0.4390 for EX, 0.4680 for HE, 0.5070 for SE, and 0.3980 for MA, achieving a mAP@50 score of 0.4530 for segmentation on the DDR validation set. When evaluated on the test set, the performance metrics were 0.4080 for EX, 0.3570 for HE, 0.2350 for SE, and 0.2080 for MA, culminating in an mAP@50 of 0.3020.

Despite the advancements made, challenges still persist, especially in detecting MA, which remains one of the most challenging lesions to identify. Additionally, the model showed lower performance in segmenting SE, which may be attributed to the limited number of examples and the lower occurrence of this lesion in the macular region.

For future work, we suggest exploring new network architectures and refining the pre-processing pipeline, including the implementation of additional data augmentation techniques during training. This could enhance the model's robustness and improve the representation of the lesions. Furthermore, we propose applying this methodology to other datasets related to diabetic retinopathy to evaluate its generalization capabilities and adaptability to various imaging conditions.

Acknowledgements

This work is partly financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) Finance Code 001. This study was partially financed by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - Brazil.

References

Alyoubi, W. L., Abulkhair, M. F., and Shalash, W. M. (2021). Diabetic retinopathy fundus image classification and lesions localization system using deep learning. *Sensors*, 21(11).

Elloumi, Y. and Kachouri, R. (2023). A robustness study of machine learning based methods for macula detection in pathological fundus images. In *2023 Twelfth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6.

Jocher, G., Chaurasia, A., and Qiu, J. (2023). Ultralytics yolov8.

Jocher, G. and Qiu, J. (2024). Ultralytics yolo11.

Kumar, N. S. and Ramaswamy Karthikeyan, B. (2021). Diabetic retinopathy detection using cnn, transformer and mlp based architectures. In *Proc. of the 2021 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, pages 1–2, Hualien , Taiwan. IEEE.

Li, T., Gao, Y., Wang, K., Guo, S., Liu, H., and Kang, H. (2019). Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening. *Information Sciences*, 501:511–522.

Munuera-Gifre, E., Saez, M., Juvinyà-Canals, D., Rodríguez-Poncelas, A., Barrot-de la–Puente, J.-F., Franch-Nadal, J., Romero-Aroca, P., Barceló, M. A., and Coll-de Tuero, G. (2020). Analysis of the location of retinal lesions in central retinographies of patients with type 2 diabetes. *Acta Ophthalmologica*, 98(1):e13–e21.

Murugan, R. and Roy, P. (2022). Micronet: microaneurysm detection in retinal fundus images using convolutional neural network. *Soft Computing: A Fusion of Foundations, Methodologies and Applications*, 26(3):1057–1066.

of Ophthalmology, I. C. (2017). Updated 2017 ico guidelines for diabetic eye care. *ICO Guidelines for Diabetic Eye Care*, pages 1–33.

Ometto, G., Assheton, P., Calivá, F., Chudzik, P., Al-diri, B., Hunter, A., and Bek, T. (2017). Spatial distribution of early red lesions is a risk factor for development of vision-threatening diabetic retinopathy. *Diabetologia*, 60(12):2361–2367.

Pereira, A., Santos, C., Aguiar, M., Welfer, D., Dias, M., and Ribeiro, M. (2023). Improved detection of fundus lesions using yolor-csp architecture and slicing aided hyper inference. *IEEE Latin America Transactions*, 21(7):806–813.

Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabuddhe, V., and Meriaudeau, F. (2018). Indian diabetic retinopathy image dataset (idrid): A database for diabetic retinopathy screening research. *Data*, 3(3).

Santos, C., Aguiar, M., Welfer, D., and Belloni, B. (2022). A new approach for detecting fundus lesions using image processing and deep neural network architecture based on yolo model. *Sensors*, 22(17).

Santos, C., Aguiar, M., Welfer, D., Dias, M., Pereira, A., Ribeiro, M., and Belloni, B. (2023). A new approach for fundus lesions instance segmentation based on mask r-cnn x101-fpn pre-trained architecture. *IEEE Access*, 11:43603–43618.

Santos, C., De Aguiar, M. S., Welfer, D., and Belloni, B. (2021). Deep neural network model based on one-stage detector for identifying fundus lesions. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, Shenzhen, China, 18–22 July 2021.

Tian, Y., Ye, Q., and Doermann, D. (2025). Yolov12: Attention-centric real-time object detectors.

Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., and Ding, G. (2024a). Yolov10: Real-time end-to-end object detection. *arXiv preprint arXiv:2405.14458*.

Wang, C.-Y., Yeh, I.-H., and Liao, H.-Y. M. (2021). You only learn one representation: Unified network for multiple tasks.

Wang, C.-Y., Yeh, I.-H., and Liao, H.-Y. M. (2024b). Yolov9: Learning what you want to learn using programmable gradient information.

Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., and Girshick, R. (2019). Detectron2. `https://github.com/facebookresearch/detectron2`.