Improving Diabetic Retinopathy Lesion Detection and Segmentation via a Macula-Guided Region of Interest Pipeline Based on YOLOv12

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 leading cause of vision impairment worldwide and is diagnosed through retinal lesions such as hard exudates, hemorrhages, soft exudates, and microaneurysms. Among these, microaneurysms represent the earliest clinical manifestation but remain difficult to detect due to their small size and low contrast. This study proposes a macula-guided Region of Interest (ROI) pipeline to improve retinal lesion detection and segmentation by exploiting the anatomical relevance of the macular region, where lesions are often more concentrated. The proposed framework employs two YOLOv12-based models: the first detects the optic disc and macula to extract ROIs, while the second detects and segments lesions using these regions during training. Experiments conducted on the Dataset for Diabetic Retinopathy achieved an mAP@50 of 0.3240 for detection and 0.3340 for segmentation. To assess generalization, the model was further evaluated on the Indian Diabetic Retinopathy Image Dataset, obtaining an mAP@50 of 0.3990 for detection and 0.2910 for segmentation. The results indicate that anatomically guided region-based training may improve predictive performance, particularly for subtle and small-scale lesions.*

1. Introduction

Diabetes is a global public health challenge characterized by insufficient insulin production or an inability to use it effectively [Martins et al. 2021]. Among the associated complications, diabetic retinopathy (DR) is a prominent microvascular disease

involving progressive damage to retinal blood vessels [Alyoubi et al. 2021]. DR affects approximately one-third of diabetic patients and is a leading cause of vision loss [International Council of Ophthalmology 2017].

Clinical diagnosis of DR involves analyzing fundus images to identify characteristic lesions, including hard exudates (EX), hemorrhages (HE), soft exudates (SE), and microaneurysms (MA). MAs represent the earliest clinical manifestations of the disease; however, their identification remains challenging due to their small size.

Prior research [Dias et al. 2025, Dias et al. 2026] investigated the utilization of the macula as a Region of Interest (ROI), demonstrating that this strategy enhances retinal lesion detection. The macular region often exhibits a higher lesion density, especially MA lesions, making it highly relevant for model training. Consequently, this study proposes a pipeline for lesion detection and segmentation. The approach integrates preprocessing and automated ROI extraction from the macular region and trains models directly on these areas, focusing the learning process on clinically significant regions.

The primary contributions are: (1) the development of a unified preprocessing workflow that integrates anatomical localization and ROI extraction into a single pipeline for retinal lesion detection and segmentation, reducing processing redundancy while improving consistency across all processing stages; (2) the enhancement of optic disc and macula detection, enabling successful ROI extraction for all images and eliminating the need for sample exclusion required by previous methodologies; (3) an ROI-guided training approach designed to prioritize anatomically relevant regions during retinal lesion detection and segmentation, with particular emphasis on small and challenging lesions such as MA; and (4) an assessment of the proposed methodology under domain shift, evaluating its ability to maintain performance across distinct retinal image domains and addressing the limitations of previous studies restricted to a single dataset.

2. Related Work

Clinical studies indicate that DR lesions are not randomly distributed but predominantly occur in the temporal regions, particularly near the macula, where they have prognostic significance [Ometto et al. 2017, Munuera-Gifre et al. 2020, Romero-Aroca et al. 2024]. Specifically, [Ometto et al. 2017] report that the proximity of early-stage lesions, such as MA and HE, to the macula correlates with disease progression. These findings underscore the clinical importance of the macula and motivate the development of methodologies that prioritize specific anatomical regions to enhance feature extraction and model performance.

In this context, [Dias et al. 2025, Dias et al. 2026] investigated the utilization of the macula as an ROI, demonstrating that spatial restriction facilitates improved feature extraction and enhances performance in retinal lesion detection and segmentation. These studies reported consistent improvements, particularly for MA and EX, which are further augmented when integrated with preprocessing strategies. However, such approaches have been evaluated on a single dataset, which limits the assessment of generalizability and robustness across varied clinical scenarios. Furthermore, the need to discard images when ROI extraction fails reduces the practical applicability of these methods.

Conversely, image division-based approaches enhance performance by optimizing local feature extraction [Santos et al. 2023, Pereira et al. 2023]. These methodologies

combine preprocessing techniques, such as partial background removal and data augmentation, with image partitioning into multiple blocks or sliding-window techniques. Nevertheless, these strategies significantly increase computational overhead by multiplying the number of samples processed and including regions of limited clinical relevance.

Clinical research consistently highlights the importance of the macular region for retinal lesion analysis [Ometto et al. 2017, Munuera-Gifre et al. 2020, Romero-Aroca et al. 2024], and contemporary computational methods increasingly leverage this information [Dias et al. 2025, Dias et al. 2026]. Although previous ROI-based strategies have demonstrated promising results for retinal lesion detection and segmentation, they remain dependent on accurate macula localization and may require image exclusion when ROI extraction fails. In addition, these approaches have been evaluated primarily on a single dataset, limiting the assessment of robustness and transferability across different domains. Therefore, there is a need for more reliable anatomically guided approaches that can effectively support retinal lesion detection and segmentation while demonstrating generalization capability across distinct retinal image domains.

3. Materials and Methods

YOLOv12 constitutes a real-time detector within the YOLO family that integrates attention mechanisms to enhance feature representation while preserving computational efficiency [Tian et al. 2025]. Adopting a backbone-neck-head architecture, the model improves feature aggregation and multiscale information fusion compared with previous YOLO versions. Additionally, its attention mechanisms capture long-range spatial dependencies, providing a richer global representation and more stable training without compromising inference efficiency [Tian et al. 2025]. Figure 1 presents an overview of the proposed methodology.

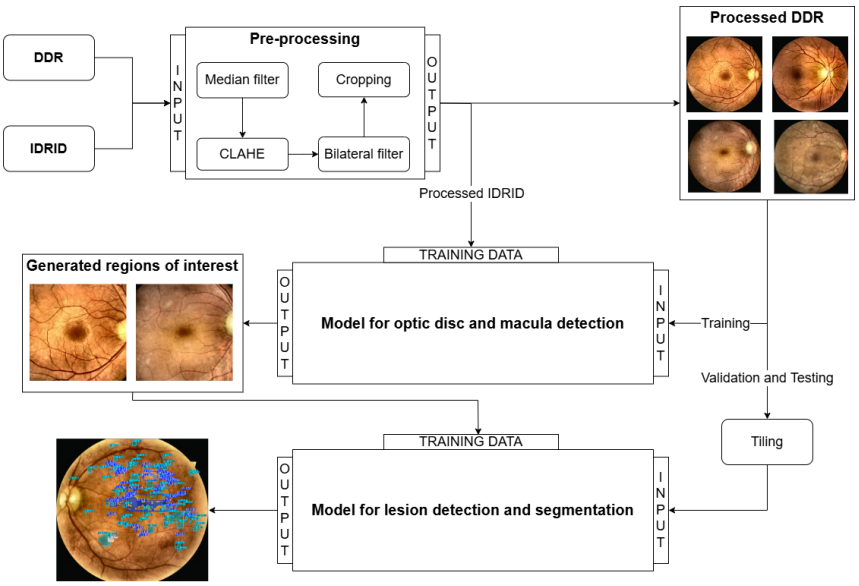


Figure 1. Overview of the proposed methodology. IDRiD images are used to train the optic disc and macula detector, enabling ROI extraction on DDR images. The extracted ROIs are subsequently used for retinal lesion detection and segmentation, while a tiling strategy is applied during evaluation. Source: Authors.

Experiments were conducted on a system equipped with an i7-10700F CPU @ 2.90GHz $\times$ 16, 16 GB of RAM, and a GeForce RTX 3060 GPU with 12 GB of VRAM. The proposed training pipeline comprises two stages: (1) ROI extraction and (2) retinal lesion detection and segmentation using the extracted ROIs. In the first stage, a YOLOv12 model is trained on the Indian Diabetic Retinopathy Image Dataset (IDRiD) [Porwal et al. 2018], which contains 516 images annotated for optic disc and macula localization. This model is subsequently applied to the Dataset for Diabetic Retinopathy (DDR) [Li et al. 2019], comprising 757 fundus images annotated for the main DR lesions, to generate macula-centered ROIs. In the second stage, these regions are used to train a YOLOv12 model for retinal lesion detection and segmentation. To assess the generalization capability of the proposed approach, the trained lesion model was evaluated on 81 IDRiD images annotated with lesions.

3.1. Performance Metrics

This study uses Average Precision (AP), defined as the area under the Precision-Recall curve. In this representation, precision quantifies the proportion of accurate predictions among all detections, while recall indicates the model's capacity to identify all positive instances [Santos et al. 2023]. Mean Average Precision (mAP) is also employed to calculate the mean AP across all categories in the dataset, enabling a comprehensive global evaluation. Equation 1 provides the formulation utilized to compute mAP, where n denotes the number of lesion categories.

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

3.2. Preprocessing

A uniform preprocessing pipeline was applied to both datasets to improve image quality and reduce background interference. First, a median filter ($kernel = 3$) was used to attenuate impulsive noise. Next, Contrast Limited Adaptive Histogram Equalization (CLAHE) ($clipLimit = 4, tileGridSize = (8, 8)$) was applied to the luminance channel in the LAB color space to enhance local contrast while preserving the original color information. A bilateral filter ($d = 9$, $\sigma_{color} = 75$, $\sigma_{space} = 75$) was then employed to smooth the image while preserving relevant retinal structures. Finally, the circular fundus region was identified using the Hough Gradient Transform (21HT), allowing partial removal of the surrounding black background. Figure 2 illustrates the adopted preprocessing workflow.

3.3. Extraction of the Region of Interest

ROI extraction relies on detecting the optic disc and macula using the IDRiD dataset, which provides annotations for both anatomical structures. Although the ROI is centered on the macula, the optic disc serves as an anatomical reference for estimating its location when direct detection fails. The dataset was divided into training, validation, and test sets in proportions of 50%, 20%, and 30%, respectively. YOLOv12x was adopted for anatomical structure detection, and Table 1 summarizes the hyperparameters used to train this model. Data augmentation included rotations ($\pm 15^{\circ}$), translations (up to 20%), scaling (up to 0.5), vertical and horizontal flips (0.5), and Mosaic augmentation (1.0).

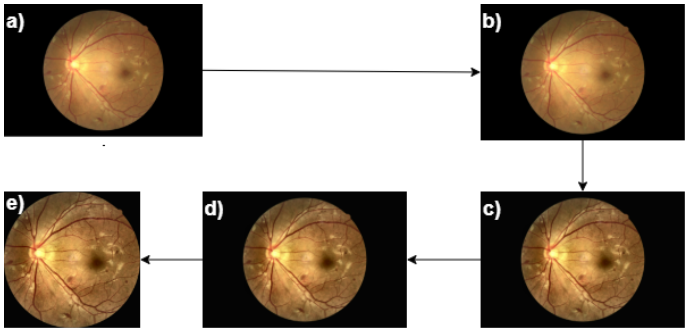


Figure 2. Preprocessing workflow applied to fundus images: (a) original image; (b) median filtering; (c) CLAHE enhancement; (d) bilateral filtering; and (e) partial black background removal. Source: Authors.

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

Hyperparameters	Value
Batch Size	8
Dropout Rate	10%
Epochs	300
Input Size	640
Learning Rate	0.001
Optimizer	AdamW
Patience	50
Seed	0

Table 2 summarizes the performance obtained on the IDRiD validation and test sets, demonstrating reliable detection of the optic disc and macula, with mAP@50 values above 0.97 in both splits. Based on these results, a final model was trained using the complete IDRiD dataset and subsequently applied to DDR for anatomical structure detection. The model achieved 97.75% accuracy for macula detection and 100% for the optic disc, surpassing the 96.40% accuracy reported in [Dias et al. 2025, Dias et al. 2026].

Table 2. Performance results for optic disc and macula detection on the IDRiD validation and test sets, utilizing the mAP@50 metric.

Dataset	Optic Disc	Macula	mAP
Validation	0.9950	0.9580	0.9760
Test	0.9900	0.9610	0.9750

Despite the high detection performance, the model was unable to directly identify the macula in 20 DDR images. To address this limitation, an anatomical property of the retina is exploited: the macula is positioned approximately 2.5 optic disc diameters from the optic disc in the temporal direction [Elloumi and Kachouri 2023]. Based on the optic disc bounding box, eye laterality is inferred, and the temporal direction is determined. The macular position is subsequently estimated by shifting from the optic disc center by approximately 2.5 times its diameter, as derived from the bounding box dimensions. Although it is based on bounding boxes rather than precise segmentation, this estimation provides an approximation of the macular location. This strategy enables macular localization for all DDR images, addressing the limitations of prior studies that discarded samples due to detection failure. Figure 3 illustrates the ROI extraction process.

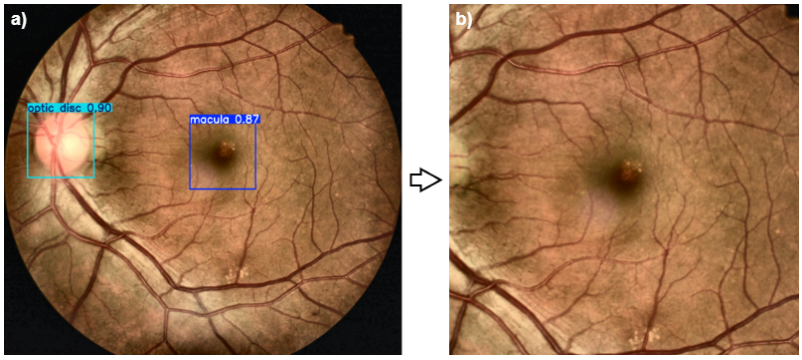


Figure 3. ROI extraction process based on anatomical localization. (a) Optic disc and macula detection; (b) ROI extraction centered on the detected or estimated macula. Source: Authors.

3.4. Training

Following the preceding stage, the model is trained to detect and segment retinal lesions using ROIs. The DDR dataset is employed for this purpose, partitioned into training, validation, and test sets in proportions of 50%, 20%, and 30%, respectively. The validation set facilitates hyperparameter optimization, whereas the test set remains isolated and is utilized exclusively for final evaluation. The study uses YOLOv12, a single-stage detector that represents one of the most recent iterations of the YOLO family.

Given that MAs are among the smallest retinal lesions and often occupy only a few pixels in fundus images, high-resolution inputs (896×896) were employed to preserve fine-grained details and improve lesion detection. However, the increased image resolution substantially raises memory consumption during training. To balance representational capacity and computational cost, YOLOv12m was selected instead of larger variants. The resulting model comprises 314 layers and 22,505,376 parameters, enabling stable training on the available hardware while maintaining sufficient capacity for retinal lesion detection and segmentation. The model was trained exclusively on the extracted ROIs for up to 400 epochs, with early stopping (patience = 70) to mitigate overfitting. A batch size of 4 was adopted due to the GPU memory requirements associated with high-resolution inputs. Table 3 summarizes the training configuration.

For the segmentation task, a mask ratio of 2 controls the downsampling of segmentation masks during training. However, due to the high supervision density introduced by mask prediction, using default loss weights causes the model to prioritize segmentation, thereby adversely affecting bounding box optimization. To rectify this imbalance, loss weights are adjusted, specifically setting the bounding box, classification, and distribution focal losses to 10, 0.8, and 1.5, respectively, thereby facilitating a more balanced optimization between detection and segmentation.

To improve model generalization and mitigate overfitting, data augmentation techniques are applied exclusively during training. The adopted transformations comprise two categories: (1) sample-mixing techniques, including Mosaic (probability = 1.0), Copy-Paste (0.3), MixUp (0.1), and CutMix (0.2); and (2) geometric transformations, including vertical and horizontal flips (probability = 0.5), rotation ($\pm 15°$), and translation (up to 20% of the image dimensions).

Table 3. Hyperparameters used for training the retinal lesion detection and segmentation model (EX, HE, SE, and MA) on the DDR dataset.

Hyperparameters	Value
Batch Size	4
Bounding Box Loss (box)	10
Classification Loss (cls)	0.8
Distribution Focal Loss (dfl)	1.5
Dropout Rate	10%
Epochs	400
Input Size	896
Learning Rate	0.00125
Mask Ratio	2
Momentum	0.937
Optimizer	AdamW
Patience	70
Weight Decay	0.001
Seed	0

3.5. Evaluation

During training, ROIs are used to focus training on anatomically relevant regions where lesions are frequently reported in clinical studies, particularly MA. Although this strategy facilitates the acquisition of relevant patterns, it restricts access to the global image context. The use of ROIs during evaluation may introduce bias, as lesions may lie outside these predefined regions. To address this limitation, a tiling strategy partitions each image into smaller patches, enabling lesion detection across the entire fundus. This approach further enhances the detection of small-scale objects by increasing their relative resolution within each patch. Consequently, all lesions are incorporated into the evaluation process, ensuring a more comprehensive analysis. Figure 4 illustrates the comparison between the original image and the generated patches.

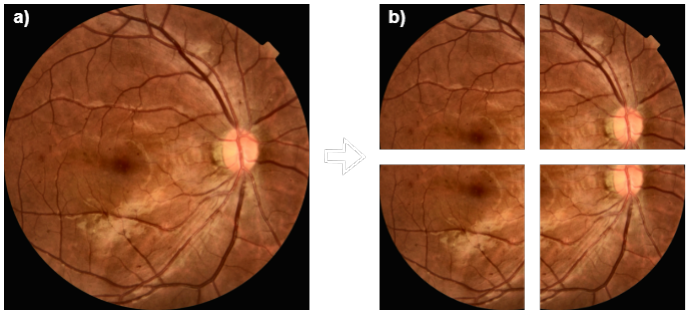


Figure 4. Tiling strategy adopted during evaluation. (a) Original fundus image. (b) Image partitioned into patches for full-image analysis. Source: Authors.

4. Results

This study compares YOLOv8 [Jocher et al. 2023], YOLOv9 [Wang et al. 2024], YOLOv11 [Jocher and Qiu 2024], and YOLOv12 [Tian et al. 2025] as state-of-the-art single-stage detectors. We initialized all models with pretrained weights and trained them using their default YOLO configurations, selecting deeper variants for comparison. We also compare the results with related studies using YOLOv12

[Dias et al. 2025], YOLOv11 [Dias et al. 2026], Detectron2 [Santos et al. 2023], and YOLOR-CSP [Pereira et al. 2023].

Table 4 presents the findings for the detection (bbox) and segmentation (mask) of EX, HE, SE, and MA lesions utilizing AP@50 and mAP@50 on the test set, which remained isolated throughout the training process. Evaluation on this dataset is essential, as it provides a more robust estimate of model generalization than the validation set used for hyperparameter optimization. The highlighted values indicate the highest performance obtained among the evaluated models for each lesion category.

Table 4. Performance results for detection (BBox) and segmentation (Mask) achieved by the proposed methodology on the DDR dataset, utilizing the AP and mAP@50 metrics.

Model		AP				mAP
		EX	**HE**	**SE**	**MA**	
YOLOv8 [Jocher et al. 2023]	BBox	0.1700	0.1600	0.1930	0.0309	0.1390
	Mask	0.1440	0.1180	0.1870	0.0240	0.1180
YOLOv9 [Wang et al. 2024]	BBox	0.1560	0.1560	0.1900	0.0271	0.1320
	Mask	0.1340	0.1110	0.1700	0.0220	0.1090
YOLOv11 [Jocher and Qiu 2024]	BBox	0.1610	0.1440	0.1620	0.0316	0.1250
	Mask	0.1400	0.1040	0.1680	0.0228	0.1080
YOLOv12 [Tian et al. 2025]	BBox	0.1540	0.1520	0.1710	0.0286	0.1260
	Mask	0.1350	0.1110	0.1550	0.0263	0.1070
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	–	–	–	–	–
YOLOv12 [Dias et al. 2025]	BBox	0.3960	0.3680	0.2280	0.1770	0.2920
	Mask	0.4080	0.3570	0.2350	0.2080	0.3020
YOLOv11 [Dias et al. 2026]	BBox	0.4090	0.3410	0.1760	0.1830	0.2770
	Mask	0.4230	0.3300	0.1780	0.2010	0.2830
Proposed approach	BBox	**0.4270**	**0.3870**	**0.2780**	**0.2030**	**0.3240**
	Mask	**0.4580**	**0.3750**	**0.2710**	**0.2300**	**0.3340**

The results presented in Table 4 indicate promising performance for retinal lesion detection and segmentation. In the detection task, the proposed model achieves AP values of 0.4270, 0.3870, 0.2780, and 0.2030 for EX, HE, SE, and MA, respectively, with an mAP@50 of 0.3240, achieving the highest mAP@50 and AP values across all lesion categories. Regarding segmentation, a similar trend is observed, with AP values of 0.4580, 0.3750, 0.2710, and 0.2300, and an mAP@50 of 0.3340, indicating slightly superior performance compared to detection.

Results on the DDR test set indicate that the proposed ROI-guided pipeline achieves superior performance compared with the evaluated reference methodologies. Notably, the model achieves promising results for MA detection, which remains challenging due to the lesion's small size and low contrast. Whereas the top-performing reference models achieve 0.1830 in detection and 0.2080 in segmentation, the proposed model attains 0.2030 and 0.2300, respectively, corresponding to improvements of approximately 10.9% and 10.6%. Even modest enhancements in MA detection remain relevant, given the intrinsic difficulty of this lesion.

Overall, the proposed approach achieves improved performance across all lesion classes, suggesting that prioritizing the macular region may contribute to more effective detection and segmentation. The approach achieves higher AP values for EX and HE, while particularly significant gains occur for more challenging lesions, such as SE and,

especially, MA. The enhanced segmentation performance further suggests more precise delineation of the affected regions. These findings suggest that focusing on the macular region may help guide the model toward clinically relevant areas, enhancing the detection of small, low-contrast lesions and placing greater emphasis on MA identification. Figure 5 illustrates an example of model prediction on the DDR test set.

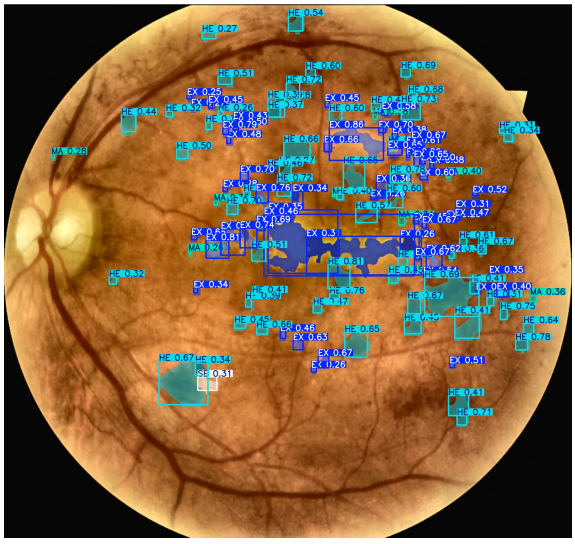


Figure 5. Illustration of model predictions for retinal lesion detection and segmentation within an image from the DDR test set. The identified lesion categories are EX (hard exudates), HE (hemorrhages), SE (soft exudates), and MA (microaneurysms). Source: Authors.

To evaluate generalization across datasets, the proposed lesion detection and segmentation model, trained exclusively on the DDR dataset, was evaluated on the IDRiD dataset. This protocol assesses the model under a distinct domain characterized by differences in image acquisition conditions, lesion distribution, and image quality. Although IDRiD was used to train the auxiliary anatomical localization model for optic disc and macula detection, this model was not used to train the lesion model. The lesion model was trained exclusively on DDR and directly evaluated on IDRiD using the tiling strategy, without using IDRiD lesion annotations during training. This protocol provides a more rigorous assessment of the model's ability to transfer across retinal image domains.

A direct comparison with related works is not straightforward, as studies such as [Santos et al. 2023, Pereira et al. 2023] train models specifically on the IDRiD dataset. In contrast, the current experiment evaluates the transferability of a model trained in a different domain. Similarly, the works in [Dias et al. 2026, Dias et al. 2025] conduct experiments exclusively on the DDR dataset. Therefore, only baseline models are considered as references for comparison in this cross-dataset evaluation setting. Table 5 presents the results obtained for the detection (bbox) and segmentation (mask) of EX, HE, SE, and MA lesions on the IDRiD dataset, utilizing AP@50 and mAP@50 as evaluation metrics. The highlighted values indicate the highest performance obtained among the evaluated models for each lesion category.

The results in Table 5 indicate that the proposed methodology achieved superior performance in both tasks within a cross-domain generalization context. For detection,

Table 5. Performance results for detection (BBox) and segmentation (Mask) achieved by the proposed methodology on the IDRiD dataset, utilizing the AP and mAP@50 metrics.

Model		AP				mAP
		EX	**HE**	**SE**	**MA**	
YOLOv8 [Jocher et al. 2023]	BBox	0.1740	0.2740	0.3870	0.0670	0.2260
	Mask	0.1240	0.1140	0.3020	0.0566	0.1490
YOLOv9 [Wang et al. 2024]	BBox	0.1570	0.2760	**0.4560**	0.0641	0.2380
	Mask	0.1170	0.1280	0.3360	0.0516	0.1580
YOLOv11 [Jocher and Qiu 2024]	BBox	0.1680	0.2760	0.4060	0.0866	0.2340
	Mask	0.1090	0.1340	0.3620	0.0615	0.1670
YOLOv12 [Tian et al. 2025]	BBox	0.1460	0.2830	0.4380	0.0748	0.2350
	Mask	0.1170	0.1380	0.3820	0.0698	0.1770
Proposed approach	BBox	**0.4130**	**0.3540**	0.4339	**0.3940**	**0.3990**
	Mask	**0.2770**	**0.2560**	**0.4040**	**0.2290**	**0.2910**

the model attains AP values of 0.4130, 0.3540, 0.4339, and 0.3940 for EX, HE, SE, and MA, respectively, with the highest mAP@50 of 0.3990, surpassing the evaluated models in all categories except SE. For segmentation, the proposed model achieves the highest performance across all lesion classes, with AP values of 0.2770, 0.2560, 0.4040, and 0.2290 for EX, HE, SE, and MA, respectively, and an mAP@50 of 0.2910.

The proposed methodology outperforms the reference models across most lesion categories, with a notable improvement in MA, one of the most challenging lesion types. In detection, baseline models achieve relatively low performance, with the highest AP reaching 0.0866, whereas the proposed model attains 0.3940, representing a clear improvement. In segmentation, the proposed model also outperforms the reference methods, indicating consistent performance across both tasks. These findings suggest that, under cross-domain generalization, the proposed methodology shows promising performance in identifying small, low-contrast lesions such as MA.

Furthermore, a comparison with the results obtained on the DDR dataset (Table 4) shows that, despite domain differences, the model maintains competitive performance across both datasets, with higher detection mAP@50 and a moderate decrease in segmentation mAP@50. This relative consistency provides evidence of generalization under domain shift and suggests potential transferability across retinal image domains.

5. Conclusion

This study proposes a retinal lesion detection and segmentation pipeline that uses the macula as an ROI to guide model training toward clinically relevant areas. The methodology integrates preprocessing, automatic ROI extraction, and YOLOv12-based retinal lesion detection and segmentation into a single workflow. In contrast to prior research, the proposed methodology unifies preprocessing and improves anatomical localization, enabling ROI extraction for all images without sample exclusion. Additionally, the study extends previous ROI-based investigations by evaluating the proposed methodology across distinct retinal image datasets. These contributions resulted in improved retinal lesion detection and segmentation performance, with particularly notable gains for MA.

In experiments on DDR, the proposed model outperformed reference methodologies across both tasks, achieving mAP@50 values of 0.3240 for detection and 0.3340 for segmentation. Particular emphasis is placed on MA, for which the model attained AP

values of 0.2030 in detection and 0.2300 in segmentation, representing a meaningful relative improvement over baseline methods, although the absolute AP values remain modest, highlighting the intrinsic difficulty of detecting this lesion type.

Generalization capability was evaluated by applying the model trained on DDR to the IDRiD dataset. The proposed methodology achieved an mAP@50 of 0.3990 for detection and 0.2910 for segmentation, outperforming the evaluated baseline models in this cross-dataset setting. Performance remained relatively consistent with that obtained on DDR, suggesting that the proposed methodology may retain its effectiveness under domain shift. However, the relatively small number of IDRiD images with lesion annotations limits the statistical confidence of this cross-dataset evaluation. These results suggest potential transferability across retinal image domains, although further evaluation on larger and more diverse datasets is required.

Despite these advances, MA remains among the most challenging lesion categories due to its small size and low contrast, indicating that further improvements are still needed in detecting subtle retinal lesions. Nevertheless, MA is among the earliest clinical signs of DR and plays a fundamental role in early diagnosis, underscoring the need for more effective detection methods.

Future research should investigate more robust architectures, such as encoder-decoder and transformer-based models, to better capture global context and complex spatial relationships, potentially enhancing the detection of subtle lesions. Additionally, the preprocessing pipeline may be refined through techniques such as super-resolution to enhance low-resolution images and reveal fine details. Future studies should also explore advanced data augmentation, class-balancing strategies, and evaluate the proposed methodology on additional datasets.

Declaration on the Use of Generative AI

Generative AI tools, such as Gemini and Grammarly, were used to refine the language during preparation. The authors did not fabricate data, alter experimental results, create references, or misrepresent their scientific interpretation. They reviewed and verified all content and take full responsibility for the manuscript's final version.

Acknowledgements

This work is partly financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), Finance Code 001.

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).

Dias, M., Santos, C., Pereira, A., Aguiar, M., and Welfer, D. (2025). Enhanced detection and segmentation of diabetic retinopathy lesions using a region of interest strategy. In *Anais do LII Seminário Integrado de Software e Hardware*, pages 61–72, Porto Alegre, RS, Brasil. SBC.

Dias, M., Santos, C., Pereira, A., Aguiar, M., and Welfer, D. (2026). Retinal lesion detection and segmentation using a region of interest-based approach with yolo11. In de Freitas, R. and Furtado, D., editors, *Intelligent Systems*, pages 440–454, Cham. Springer Nature Switzerland.

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.

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

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

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

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.

Martins, J., Coelho, A., and Baffa, M. (2021). Detecção dos estágios iniciais da retinopatia diabética utilizando redes neurais profundas. In *Proc. of the VIII Escola Regional de Computação Aplicada à Saúde (ERCAS 2021)*, pages 42–45, Porto Alegre, RS, Brasil. SBC.

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.

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).

Romero-Aroca, P., Garcia-Curto, E., Pascual-Fontanilles, J., Valls, A., Moreno, A., and Baget-Bernaldiz, M. (2024). Distribution of microaneurysms and hemorrhages in accordance with the grading of diabetic retinopathy in type diabetes patients. *Diagnostics*, 14(14).

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.

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

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