YOLO11 and YOLO12 Deep Learning Models for Diabetic Retinopathy Classification: A Comparative Analysis

1st Maria C. A. Castro
Teleinformatics Engineering Department (DETI)
Federal University of Ceara (UFC)
Fortaleza, Brazil
cecialves@alu.ufc.br

2nd Carlos V. G. Moura
Teleinformatics Engineering Department (DETI)
Federal University of Ceara (UFC)
Fortaleza, Brazil
carlos.victor@alu.ufc.br

3rd Débora F. de Assis
Teleinformatics Engineering Department (DETI)
Federal University of Ceara (UFC)
Fortaleza, Brazil
debora.ferreira@lesc.ufc.br

4th Paulo C. Cortez
Teleinformatics Engineering Department (DETI)
Federal University of Ceara (UFC)
Fortaleza, Brazil
cortez@lesc.ufc.br

***Abstract*—Diabetic retinopathy (DR) is a progressive microvascular complication of diabetes mellitus that affects the retina, causing changes in the blood vessels. In its early stages, the disease is asymptomatic, but as microvascular changes progress, the patient may experience impaired vision. The disease can be detected early through a fundus examination, in which the ophthalmologist looks for characteristic lesions in the image. Professionals often have difficulty in diagnosis due to the high volume of exams and the small size of the lesions in the image. DR can be prevented with glycemic control and regular eye exams, and early detection is essential to preserve the patient's vision. Computer vision techniques can assist professionals in the diagnosis of DR through automatic image classification of the disease. This study aims to develop a multi-class classification method to identify the severity level of DR. To this end, the DDR and IDRiD dataset were used to train and validate the classification models. The YOLO11 and YOLO12 base models were trained in this task, and YOLO11 was also trained starting with pre-trained weights. A class reduction approach was also used, reducing the number of classes from 5 to 3 by merging classes from similar DR severity levels. The pre-trained YOLO11 model obtained the best results, achieving 71.04% mean accuracy for 5-class classification and 86.15% for 3-class. The model also achieved 88.33%¨accuracy classifying the earlier stages of DR. Thus, the proposed method has the potential to aid ophthalmologists in the clinical diagnosis of DR, contributing an efficient and accessible approach to DR arge-scale screening.**

***Index Terms*—diabetic retinopathy, image classification, YOLO, deep learning**

I. Introduction

Diabetic Retinopathy (DR) is a complication of diabetes which can compromise vision and is often associated with other serious health issues. The progression of DR indicates a high risk of micro and macrovascular diseases, such as kidney problems, strokes and cardiovascular complications [1].

In early stages, DR is asymptomatic and usually causes no noticeable visual changes. However, in more advanced stages, DR can compromise the patient's vision and even lead to loss of sight. According to studies carried out in several countries, when diagnosis and treatment are carried out early, the risk of blindness associated with DR can be reduced to less than 5% [2]. Moreover, according to the World Health Organization, around 285 million people live with some form of visual impairment, even though around 60% to 80% of these cases could have been prevented or treated. In Brazil, specially, the 2010 Demographic Census (IBGE) revealed that more than 35 million people have some degree of visual impairment, from which the main preventable cause of blindness in adults is diabetic retinopathy [3,4]. In addition, according to the IDF 2021 atlas, around 15.7 million people live with diabetes in Brazil, making it the sixth country with the highest number of people with diabetes [5]. Thus, regular ophtalmologic monitoring of people with diabetes is essential.

According to the National Commission for the Incorporation of Technologies into the Unified Health System (CONITEC), 2021, diabetic retinopathy can be classified into different stages according to the changes identified in the fundus examination with pupil dilation. In early cases, there are no visible changes. However, with the appearance of microaneurysms, the mild form of non-proliferative retinopathy (NPDR) is identified. As it progresses, hemorrhages and vessel alterations appear, defining the moderate and severe stages. When there are multiple signs combined, NPDR is considered very severe. The proliferative form (PDR) is marked by abnormal blood vessel growth and deeper hemorrhages, increasing the risk of

Authors would like to thank the Brazilian National Council for Scientific and Technological Development (CNPq) and Coordination of Superior Level Staff Improvement (CAPES)-Finance Code 001 for the support to this research.

visual loss [6].

Fundus imaging using retinal photography, fundoscopy, is the standard method for detecting and monitoring DR. The diagnosis, however, requires a specialized ophtalmologist, and has inter-observer variability. In addition, health centers often face high costs and delays in issuing reports, which limits large-scale screening campaigns.

Despite advancements in screening methods, particularly through the use of computer vision to support the diagnosis of ocular diseases, several challenges remain. Among these are the limited availability of publicly labeled datasets and the difficulty in identifying patterns indicative of Diabetic Retinopathy, as well as in detecting small lesions such as microaneurysms (MAs), hemorrhages (HEMs), and exudates (EXs).

To overcome these barriers, recent studies have explored the use of deep learning techniques applied to the automated screening of Diabetic Retinopathy (DR). With advancements in real-time classification and object detection models, the *You Only Look Once* (YOLO) series has stood out for its efficiency in automatically identifying complex visual patterns.

In this study, we propose a comparative approach of versions 11 and 12 of YOLO, which present architectural evolution with substantial improvements in stability, computational efficiency, and generalization capability. We also compare the base YOLO11 with its pre-trained in the COCO dataset version. Moreover, a DR severity class reduction approach is tested. Thus, our objective is to classify the severity of Diabetic Retinopathy from fundus images. This integration represents a significant advancement in the development of automated screening systems, with the potential for large-scale application in both clinical and community settings.

II. Related Works

Given the increasing prevalence of Diabetic Retinopathy and the inherent limitations of traditional clinical diagnosis, such as dependence on specialists, interobserver variability, and difficulty in identifying subtle lesions, numerous studies have explored the use of Artificial Intelligence as a supporting tool for screening and classifying the disease. Research in this area has shown promising results, both in the automatic detection of specific lesions and in binary or multiclass classification of DR severity, contributing to faster, more consistent, and more accessible diagnoses, particularly in regions with limited specialized resources.

Among the most notable approaches is the use of the YOLO model, which has demonstrated good performance in detecting microaneurysms, exudates, and hemorrhages, even without intensive image preprocessing [9]. Complementarily, other studies have employed traditional machine learning methods with manual feature extraction based on segmentations, highlighting the relevance of integrating segmentation, extraction, and classification in the clinical context.

Alyoubi *et al.* [1] proposed a hybrid approach that combines CNN512 and YOLOv3 models. After preprocessing steps (CLAHE, color normalization, noise removal, and data augmentation), CNN512 classified DR stages with an accuracy of 88.6% (DDR) and 84% (APTOS). YOLOv3 achieved a mAP of 0.216 in lesion detection, and its fusion with the classifier increased performance to 89% accuracy, 89% sensitivity, and 97.3% specificity, outperforming previous methods.

Other studies reinforce the versatility of the YOLO versions in DR detection. Han *et al.* [4] used YOLOv5 with normalization and data augmentation techniques on the DDR dataset to detect small DR lesions, achieving a mAP of 34%, mAUC of 0.76, and mIoU of 0.53, demonstrating the method's effectiveness for real-time clinical applications. Santos *et al.* [11] also used the public DDR dataset to detect lesions associated with diabetic retinopathy using YOLOv5, but applied CLAHE, background cropping, and data augmentation (MixUp, Mosaic, and Tiling), achieving a mAP of 0.25 and an F1-score of 0.35 in validation, with notable performance in detecting exudates and hemorrhages.

Wahab Sait [13] developed a lightweight architecture combining YOLOv7 for feature extraction, the QMPA algorithm for feature selection, and the MobileNet V3–Small classifier. Evaluated on the APTOS and EyePacs datasets, the model achieved accuracy above 98% and F1-score above 93%, with low inference time and few parameters, proving suitable for mobile applications and resource-constrained settings.

Geetha *et al.* [3] proposed the use of YOLOv8 for detection and classification of DR stages from fundus images, incorporating optimized convolutions, zero padding, and stride adjustments. The results showed stable loss curves and promising performance, with a mAP of 0.5, F1-score up to 0.59, and accuracy of 67.72

et al. [**?**] proposed an embedded DR detection system using YOLOv8 on a Jetson Nano device. With real-time inference (2.6 ms), the model achieved 93.77% training accuracy and 72.6% validation accuracy, confirming the feasibility of lightweight solutions for automated screening in infrastructure-constrained environments.

Danko *et al.* [2] trained five variations of YOLOv8n with the SAHI algorithm in the DDR dataset, focusing on small object detection. Techniques such as CLAHE and dark border cropping were applied, and results indicated significant improvements, with a relative increase of up to 30.9% in IoU.

Thus, the studies discussed highlight the potential of the YOLO model, in its various versions, as an effective technique for assisted diagnosis of Diabetic Retinopathy. Motivated by these advances and the gap in comparative studies between recent YOLO architecture versions specifically aimed at DR severity classification, this work proposes an in-depth analysis of YOLO11 and YOLO12 versions applied to this task, with the goal of identifying the most suitable model for a high-precision automated screening system.

III. Materials and Methods

For this study, color fundus images from the public Diabetic Retinopathy Database (DDR) were used for training and validation. To ensure compatibility with the input requirements

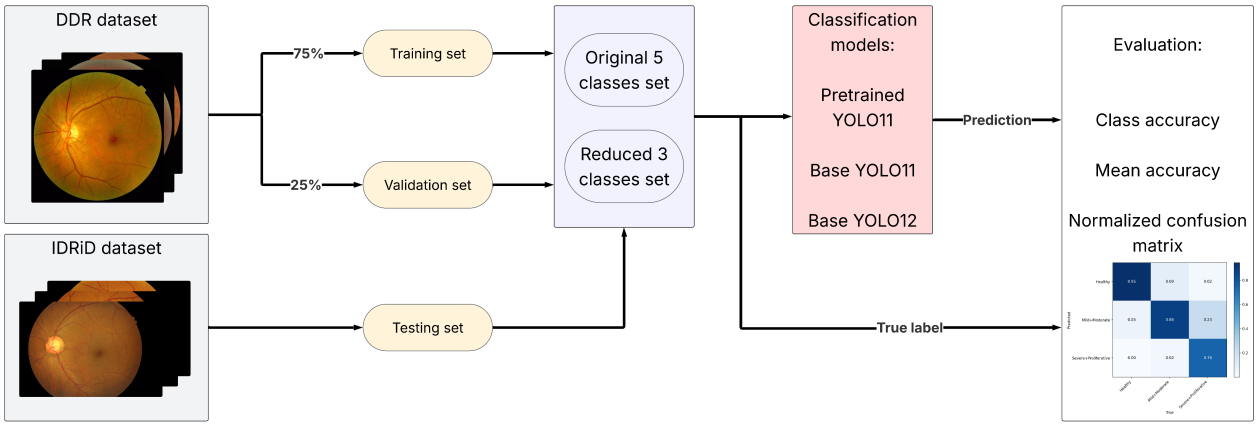

Fig. 1: Diagram of overall workflow process.

of the YOLO models, all images underwent a preprocessing stage that included resizing (640×640) and normalization. In addition, standard data augmentation techniques, integrated into the YOLO framework, were applied to increase dataset variability and reduce the risk of overfitting.

Two versions of the YOLO model — 11 and 12 — were evaluated for the multiclass classification task, enabling the prediction of five severity levels of diabetic retinopathy: no DR, mild, moderate, severe, and proliferative. Training parameters such as batch size, number of workers, and number of epochs were tuned to 8, 4 and 50 respectively. The Adam optimizer was employed to update the network weights during training, with an automatic learning rate adjustment at each epoch based on the gradient behavior.

To address the unbalanced number of images in each DR severity class, this work proposed a class reduction approach to reduce the imbalance. Two types of models were trained: 5-class models using all 5 original DR severity classes, as well as 3-class models which merged mild and moderate classes in a single class, doing the same to the severe and proliferative classes. Images with close DR severity could show many common structures and other visual similarities [14]. For this reason, class reduction could be a viable approach to reduce class imbalance without much loss of purpose to the model.

Ultralytics provides YOLO models trained on the COCO dataset in addition to the base weight models. However, in the time of writing, there is no pretrained YOLO12 classification model in the COCO dataset. Transfer learning could be a viable approach to improve model performance in medical image classification [7]. For this reason, this work trained the pretrained YOLO11 (Pre11) model to test pretraining performance, as well as the base weight YOLO11 (Base11) and YOLO12 (Base12) models to properly compare YOLO11 and YOLO12. Pre11, Base11 and Base12 models were trained in both 5-class classification and 3-class classification, totaling 6 models trained in all.

Finally, to assess the generalization capability of the proposed models, an external test set from the Indian Diabetic Retinopathy Image Dataset (IDRiD) was employed. This dataset, which differs in terms of patient population and image acquisition conditions, enabled a more realistic and robust evaluation of model performance across diverse clinical scenarios. The methodology diagram of the method used in this work is presented in Figure 1.

A. *Datasets*

In this study, two datasets widely recognized in the literature for Diabetic Retinopathy classification tasks were used: the Diabetic Retinopathy Database (DDR) for model training and the Indian Diabetic Retinopathy Image Dataset (IDRiD) as an external validation set.

The DDR [8] is one of the largest publicly available repositories for fundus image analysis, comprising 13,673 color images collected between 2016 and 2018 from 147 hospitals in China. These images, obtained from 9,598 patients aged between 1 and 100 years (mean age of 54.13 years), are accompanied by clinical annotations that classify DR into five categories: no DR, mild, moderate, severe, and proliferative. An additional class of ungradable images was excluded from this study. In total, the dataset contains 9391 images: 4699 healthy, 473 mild, 3357 moderate, 177 severe and 685 proliferative images. A challenging aspect of this dataset is the significant class imbalance, which poses a major limitation for training multiclass classification models, as it may hinder the model's ability to correctly identify the less represented stages of the disease. DDR image samples from each DR severity level are presented in figure 2

Conversely, the IDRiD [10] dataset focuses on the analysis of DR in Indian patients and is organized into three sections: *(a) segmentation, (b) disease grading, and (c) lesion localization*. For this study, we exclusively used the disease grading section, which is dedicated to classifying the DR stage. This subset consists of 516 high-quality images, carefully validated by experts for diagnostic clarity and clinical relevance. The images are distributed among patients without DR (168 images) and with DR (326 images), the latter subdivided into the

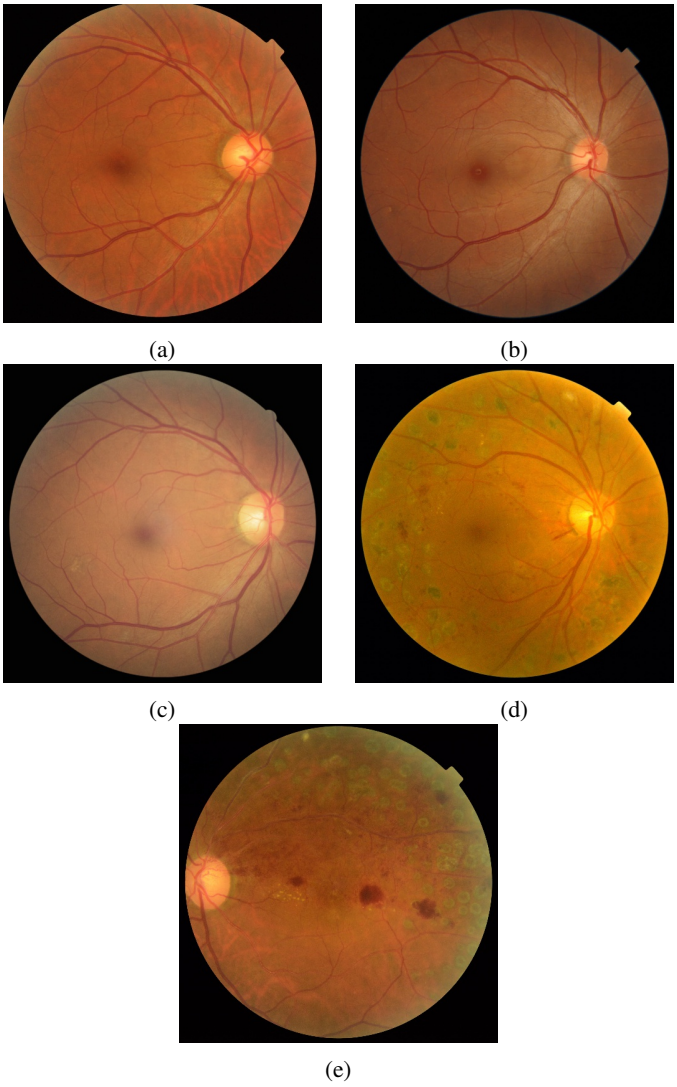

Fig. 2: Fig. 3: Sample DDR images from each DR severity level. (a) Healthy. (b) Mild. (c) Moderate. (d) Severe. (e) Proliferative.

four classical stages of the disease: 25 mild, 157 moderate, 86 severe and 58 proliferative. In this work, the IDRiD dataset was used exclusively for external validation, allowing for a broader and more realistic assessment of the generalization ability of the model trained on the DDR dataset.

In general, the combination of these two datasets enables the exploration of both the wide variability and sample representativeness of the DDR, and the diagnostic quality and standardization of the IDRiD, thus providing a robust and clinically relevant evaluation of the proposed approach.

B. YOLO11

YOLO11 [5] introduces enhancements that optimize both performance and computational efficiency, making it a robust solution for a wide range of applications and hardware capabilities, including detection, segmentation, classification, pose estimation, and oriented object detection. This version improves feature extraction through architectural refinements in the backbone and neck of the network, enabling better capture of visual details even in complex tasks. One of its main innovations is the C3k2 block, which relies on smaller convolutions while preserving feature extraction quality. Additionally, the model incorporates the C2PSA module, which introduces parallel spatial attention, allowing the model to better focus on relevant regions of the image. The SPPF module is also retained and optimized to enhance multi-scale feature aggregation. A notable highlight is the YOLO11m variant, which achieves superior results with 22% fewer parameters compared to YOLOv8m, while maintaining computational efficiency [6].

C. YOLO12

YOLO12 [12] is currently the latest version of the YOLO models family, introducing an architecture based on attention mechanisms that overcomes the traditional limitations of convolutional networks while maintaining the high speed required for real-time applications. It stands out for employing a simplified and efficient form of attention that partitions feature maps into regions to reduce computational cost, while preserving a broad receptive field. The model also introduces improvements in information aggregation through R-ELAN, which facilitates optimization in larger architectures by means of residual connections and lighter structures. Additional changes include the adoption of techniques such as FlashAttention to optimize memory access, the removal of positional encodings, the use of efficient convolutions, and adjustments in the balance between attention and feed-forward layers. These changes make YOLO12 faster and more accurate than its predecessors, outperforming models such as YOLOv10, YOLO11, and RT-DETRv2 in both accuracy and latency metrics. YOLO12 supports various tasks, including object detection, segmentation, classification, pose estimation, and oriented object detection .

D. Evaluation Metrics

To quantitatively analyze the experimental results, the accuracy (ACC) for each individual class, mean accuracy (mACC) across all classes, and the multiclass confusion matrix were used. To calculate these metrics, the images were classified as true positive (TP), true negative (TN), false positive (FP), or false negative (FN), based on the comparison between the label predicted by the model and the actual label of the image.

Accuracy represents the proportion of correct classifications within each class, enabling a more detailed analysis of the model's performance across the different stages of Diabetic Retinopathy. Mean accuracy (mACC), in turn, corresponds to the average of the individual class accuracies, providing a global measure of performance that is especially useful in scenarios involving class imbalance. The multiclass confusion matrix, on the other hand, organizes the predictions into a table that compares true labels with predicted labels. This allows for the visualization of not only correct classifications but also the specific errors made between classes. Such detailed

TABLE I: 5-class models accuracy (%)

	DDR Dataset - Validation			**IDRiD Dataset** - Test		
Class	Pre11	Base11	Base12	Pre11	Base11	Base12
Healthy	**94.26**	80.66	76.96	44.05	30.36	**51.79**
Mild	**38.22**	00.00	00.00	**32.00**	00.00	00.00
Moderate	**88.04**	77.86	79.55	**92.36**	79.62	76.43
Severe	**49.15**	00.00	00.00	**46.51**	01.16	00.00
Proliferative	**85.53**	50.88	47.37	31.03	**74.14**	65.52
mACC	**71.04**	41.88	40.78	**49.19**	37.06	38.75

TABLE II: 3-class models accuracy (%)

	DDR Dataset - Validation			**IDRiD Dataset** - Test		
Class	Pre11	Base11	Base12	Pre11	Base11	Base12
Healthy	**94.51**	79.96	77.92	38.10	33.93	**61.90**
Mild+Moderate	**88.33**	77.57	74.71	**88.14**	76.84	70.62
Severe+Proliferative	**75.61**	54.36	33.45	65.41	**63.91**	47.37
mACC	**86.15**	70.63	62.03	**63.88**	58.23	59.96

analysis reveals patterns of confusion, particularly between adjacent severity levels, and is essential for understanding the model's limitations in clinical scenarios where diagnostic precision directly influences decision-making. The formulas for computing ACC and mACC are shown in equations 1 and 2.

$$\text{ACC} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\text{mACC} = \frac{1}{K} \sum_{i=1}^{K} \text{ACC}_i \tag{2}$$

IV. Results and Discussion

Overall, the Pre11 model obtained the best results in 5-class classification in both DDR and IDRiD datasets. The model obtained 71.04% mean accuracy in all classes of the DDR dataset and 49.19% in the IDRiD dataset, as well as the highest accuracy in each individual class of the DDR dataset. However, two models obtained higher accuracies in specific classes of the IDRiD dataset: the Base12 model obtained 51.79% accuracy in the healthy class and the the Base11 model obtained 71.14% accuracy in the proliferative class of the IDRiD dataset. The accuracies of each 5-class classification models are presented in table I.

The Pre11 model's performance surpasses the 67.72% accuracy reported by Geetha *et al.* [3] obtained in multi-stage methods such as Alyoubi *et al.* [1] which combined YOLOv3 with CNN classifiers, and Wahab Sait [13], who reached over 93% F1-score on binary DR detection but with higher computational cost. Here, the same accuracy range was achieved in a single-stage, streamlined configuration, without additional preprocessing steps such as CLAHE or background cropping used in Han *et al.* [4] and Santos *et al.* [11], which reinforces the efficiency of the proposed pipeline.

The Base11 and Base12 models obtained similar results in 5-class classification. Both base models were not able to correctly identify the mild and severe classes. This could be explained by the high class imbalance of these classes and by the visual similarities between near severity classes, as the Pre11 model also incorrectly classified more than half images of these classes. It is also worth noting that both base models obtained higher accuracy in the proliferative class than the Pre11 model.

Similarly to the previous results, overall the Pre11 model obtained the best results in 3-class classification in both DDR and IDRiD datasets. The model obtained 86.15% mean accuracy in all classes of the DDR dataset and 63.88% mean accuracy in the IDRiD dataset, as well as the highest accuracy in each individual class of the DDR dataset. As in the previous results, the Base12 model also obtained the highest accuracy of 61.90% in the healthy class of the IDRiD dataset. The accuracies of each 3-class classification models are presented in table II.

The 5-class and 3-class classification models obtained similar accuracies in the healthy class of the DDR dataset. The 5-class Pre11 model obtained a higher accuracy in the healthy class of the IDRiD dataset compared to the 3-class model. However, the opposite occurred in the Base11 and Base12 models, as the 3-class models obtained a higher accuracy in the healthy class of the IDRiD dataset than their 5-class counterpart.

By merging classes with similar levels of DR severity, the model was able to predict disease stages more accurately, even across a broader range of severities. In the DDR dataset, the 3-class Pre11 model achieved 88.33% accuracy in the

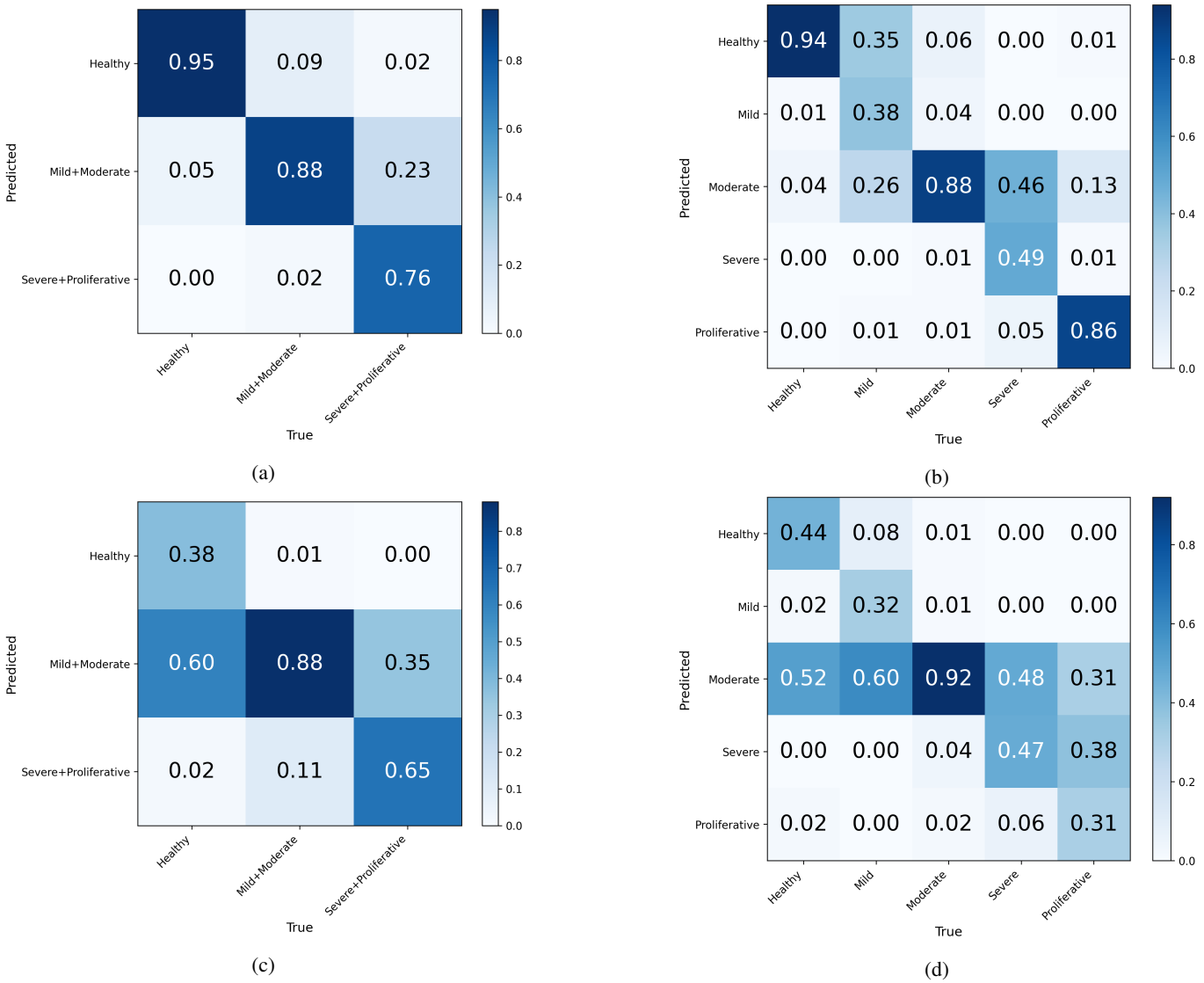

Fig. 3: Confusion matrices of each model. (a) 3-class Pre11 prediction on DDR. (B) 5-class Pre11 prediction on DDR. (c) 3-class Pre11 prediction on IDRiD. (d) 5-class Pre11 prediction on IDRiD.

mild+moderate class, outperforming the 5-class model in the individual mild and moderate classes. In the severe and proliferative classes, however, the 5-class model showed slightly better results than the 3-class model, as the latter frequently misclassified the severe+proliferative class as healthy. This trend of improved accuracy when grouping adjacent stages is consistent with observations by Danko *et al.* [2], who reported significant gains when optimizing detection for broader categories of small lesions. The normalized confusion matrices for the 3-class and 5-class Pre11 model predictions on the DDR and IDRiD datasets are shown in Figure 3.

Overall, results demonstrate that the class-merging strategy reduces ambiguities and focuses learning on more robust patterns. In medical problems like DR, where the boundaries between adjacent classes are subtle and often subjective, grouping stages with similar clinical manifestations allows the model to concentrate on separating clinically distinct categories, rather than expending capacity on distinguishing extremely similar patterns.

In the IDRiD dataset, the 3-class Pre11 model considerably improved the severe+proliferative accuracy compared to its 5-class counterpart. However, the 3-class model more frequently misclassified healthy images as DR. Overall, both 5-class and 3-class models misclassified many of images as moderate or mild+moderate in the IDRiD dataset. This could be caused by differences in the criteria used to classify images in both datasets, as well as visual variations caused by different fundoscopic instruments.

Overall, pre-training in the COCO dataset considerably improved model performance. In both 5-class and 3-class classification, Pre11 showed much better results than Base11 across both datasets. This shows the advantages that could

be gained by pretraining YOLO models for medical imaging classification, even if the images used for pretraining are not from the medical field.

For the most part, both base models obtained similar results. Despite differences in individual class accuracy, both models obtained similar mean accuracy in the IDRiD dataset for 5-class and 3-class classification, as well as in the DDR dataset for 5-class classification. However, in the 3-class classification of DDR, the Base11 model obtained 70.63% mean accuracy, while Base12 obtained 62.03%. The main improvement was also in the later stages of DR, as Base11 obtained 54.36% accuracy in the severe+proliferative class, while Base12 obtained 33.45%. This suggests that, although YOLO12 incorporates architectural advances over YOLO11, in this specific medical context the older version may retain an edge in fine-grained lesion discrimination for advanced DR stages.

V. Conclusion

This study proposes a comparative analysis between YOLO11 and YOLO12 models in the task of classifying diabetic retinopathy severity in fundus images. Model pretraining in the COCO dataset was also evaluated, and a class reduction approach was tested. The DDR public dataset was used for training and validation, and the IDRiD dataset for external testing, allowing an evaluation of the models' generalization capability in different scenarios.

Overall, the YOLO11 pre-trained model obtained the best results in both 5-class and 3-class classification, obtaining 71.04% mean accuracy in the DDR dataset and 49.19% in the IDRiD dataset for 5-class classification, as well as 86.15% in the DDR dataset and 63.88% in the IDRiD dataset for 3-class classification. The inferior performance in the IDRiD dataset compared to DDR highlights the challenges of generalizing the model for different scenarios and acquisition conditions. The class reduction approach, which merged similar severity classes in a single class, showed effective results in mitigating misclassification between visually similar classes.

By testing recently released YOLO architectures in an yet untested application, our work provides new insights to the high capabilities of YOLO11 and YOLO12 models in the diabetic retinopathy classification field. In addition, the presented method exhibits adequate performance to assist ophthalmologists in detecting and classifying diabetic retinopathy.

Future works could use additional strategies to mitigate the impact of class imbalance in model performance. Other datasets could also be used to further improve model generalization capability. In addition, the YOLO12 classification model pretrained in the COCO dataset could be tested for further comparison when it is released.

Acknowledgment

This work was made possible with the financial support of the Coordination of Superior Level Staff Improvement (CAPES) and with the support of the Laboratory of Computer Systems Engineering (LESC) - UFC.

References

[1] Wejdan L Alyoubi, Maysoon F Abulkhair, and Wafaa M Shalash. Diabetic retinopathy fundus image classification and lesions localization system using deep learning. *Sensors*, 21(11):3704, 2021.

[2] Tomáš Danko, Miloš Oravec, Veronika Kurilovrá, and Jarmila Pavlovicova. Small object detection in fundus images. In *2024 International Symposium ELMAR*, pages 105–108. IEEE, 2024.

[3] DEVI APPARI GEETHA, TV HYMA LAKSHMI, K VIDYA SAGAR, MORSA CHAITANYA, SRAVANTHI KANTAMANENI, VEERA VASANTHA RAO BATTULA, SURYA PRASADA RAO BORRA, POOJA MEENA, PRIYANKA GUPTA, DR SOHIT AGARWAL, et al. Detection and classification of diabetic retinopathy using yolo-v8 deep learning methodology. *Journal of Theoretical and Applied Information Technology*, 102(7), 2024.

[4] Zhike Han, Zeheng Ye, Ruilin Liang, Wei Dai, Hanyu Xiao, Xing Wang, Wangchao Wu, Chaoyang Hong, Qingqing Zheng, Shuiguang Deng, et al. A lightweight method for precise small lesion detection in diabetic retinopathy. *Biomedical Signal Processing and Control*, 109:108006, 2025.

[5] Glenn Jocher and Jing Qiu. Ultralytics yolo11, 2024.

[6] Rahima Khanam and Muhammad Hussain. Yolov11: An overview of the key architectural enhancements. *arXiv preprint arXiv:2410.17725*, 2024.

[7] Cosa-Linan A. Santhanam N. et al. Kim, H.E. Transfer learning for medical image classification: a literature review. *BMC Med Imaging 22, 69*, 2022.

[8] Tao Li, Yingqi Gao, Kai Wang, Song Guo, Hanruo Liu, and Hong Kang. Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening. *Information Sciences*, 501:511–522, 2019.

[9] CVG Moura, PC Cortez, DF Assis, PC Motta, and BR Silva. Yolov8 deep learning model for diabetic retinopathy fundus image segmentation and disease classification. In *XVI Brazilian Conference on Computational Intelligence (CBIC 2023), Salvador:(SBIC) Sociedade Brasileira de Inteligência Computacional*, pages 1–7, 2023.

[10] Prasanna Porwal, Samiksha Pachade, Ravi Kamble, Manesh Kokare, Girish Deshmukh, Vivek Sahasrabuddhe, and Fabrice Meriaudeau. Indian diabetic retinopathy image dataset (idrid), 2018.

[11] Carlos Santos, Marilton Aguiar, Daniel Welfer, and Bruno Belloni. A new approach for detecting fundus lesions using image processing and deep neural network architecture based on yolo model. *Sensors*, 22(17):6441, 2022.

[12] Yunjie Tian, Qixiang Ye, and David Doermann. Yolov12: Attention-centric real-time object detectors. *arXiv preprint arXiv:2502.12524*, 2025.

[13] Abdul Rahaman Wahab Sait. A lightweight diabetic retinopathy detection model using a deep-learning technique. *Diagnostics*, 13(19):3120, 2023.

[14] Keye Wong. *Defining Diabetic Retinopathy Severity*, pages 105–120. Springer New York, New York, NY, 2010.